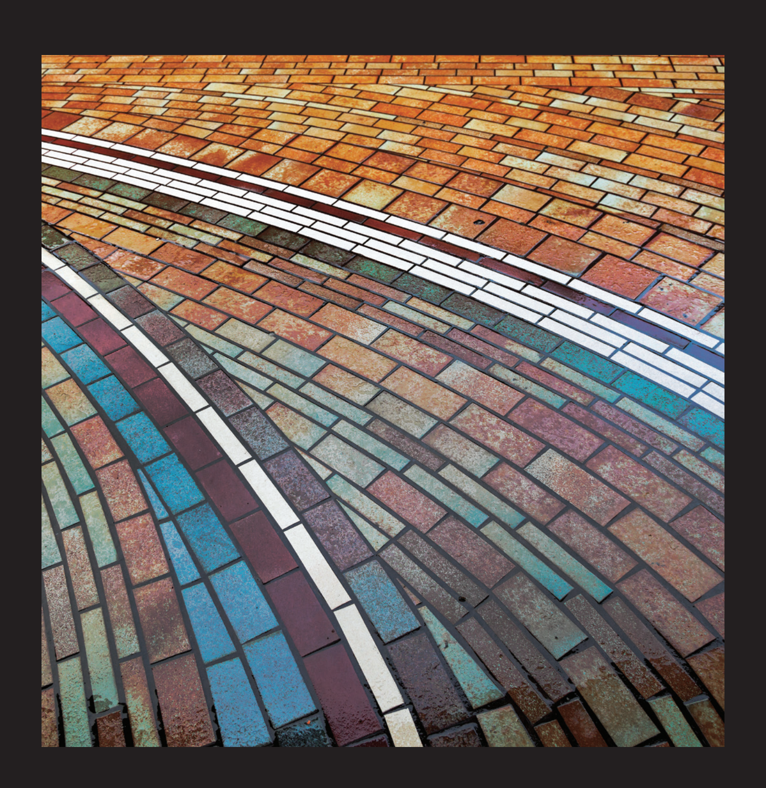
Sous la direction
de MARC CORBIÈRE et NADINE LARIVIÈRE

MÉTHODES QUALITATIVES, QUANTITATIVES ET MIXTES

Dans la recherche en sciences humaines, sociales et de la santé



Presses de l'Université du Québec

MÉTHODES QUALITATIVES, QUANTITATIVES ET MIXTES



Presses de l'Université du Québec
Le Delta I, 2875, boulevard Laurier, bureau 450, Québec (Québec) G1V 2M2
Téléphone : 418 657-4399 *Télécopieur :* 418 657-2096
Courriel : puq@puq.ca *Internet :* www.puq.ca

Diffusion / Distribution :

Canada Prologue inc., 1650, boulevard Lionel-Bertrand, Boisbriand (Québec) J7H 1N7
Tél. : 450 434-0306 / 1 800 363-2864

France AFPU-D – Association française des Presses d'université
Sodis, 128, avenue du Maréchal de Lattre de Tassigny, 77 403 Lagny, France – Tél. : 01 60 07 82 99

Belgique Patrimoine SPRL, avenue Milcamps 119, 1030 Bruxelles, Belgique – Tél. : 02 7366847

Suisse Servidis SA, Chemin des Chalets 7, 1279 Chavannes-de-Bogis, Suisse – Tél. : 022 960.95.32

Sous la direction
de MARC CORBIÈRE et NADINE LARIVIÈRE

MÉTHODES QUALITATIVES, QUANTITATIVES ET MIXTES

Dans la recherche en sciences humaines, sociales et de la santé



Catalogage avant publication de Bibliothèque et Archives nationales du Québec et Bibliothèque et Archives Canada

Vedette principale au titre :

Méthodes qualitatives, quantitatives et mixtes : dans la recherche en sciences humaines, sociales et de la santé

Comprend des références bibliographiques.

ISBN 978-2-7605-4011-8

1. Sciences sociales – Recherche. 2. Recherche qualitative.
3. Recherche quantitative. I. Corbière, Marc, 1967- .
II. Larivière, Nadine, 1970- .

H62.M47 2014 300.72 C2014-940042-X

Les Presses de l'Université du Québec
reconnaissent l'aide financière du gouvernement du Canada
par l'entremise du Fonds du livre du Canada
et du Conseil des Arts du Canada pour leurs activités d'édition.

Elles remercient également la Société de développement
des entreprises culturelles (SODEC) pour son soutien financier.

Conception graphique
Richard Hodgson

Mise en pages
Info 1000 mots

Dépôt légal : 2e trimestre 2014
› Bibliothèque et Archives nationales du Québec
› Bibliothèque et Archives Canada

Imprimé au Canada

TABLE DES MATIÈRES

CHAPITRE 18

CHAPITRE 19

CHAPITRE 23

CHAPITRE 24

LISTE DES FIGURES

LISTE DES TABLEAUX

INTRODUCTION

Marc Corbière
Nadine Larivière

Dans un monde de la recherche de plus en plus décloisonné et vaste, il devient souhaitable de disposer d'une perspective englobante, capable d'intégrer les apports de diverses approches méthodologiques, tout en reconnaissant leurs particularités propres. Il y a certes les paradigmes sous-jacents au domaine de la recherche qui orientent la conception et le déroulement d'un projet de recherche ; par exemple, une approche épistémologique chez le chercheur renvoie communément aux paradigmes constructiviste, positiviste ou néopositiviste. L'image que se fait le chercheur de la réalité peut en effet être le produit de son interaction avec cette réalité (approche constructiviste) ou encore le fruit de l'utilisation d'analyses statistiques qui lui permet de décrire, d'expliquer et de prédire des phénomènes par l'intermédiaire de variables observables et mesurables (approche positiviste), avec cependant une conception qu'il existe une possible influence du chercheur sur l'objet d'étude et que cette réalité reste de l'ordre du probable et non d'une vérité absolue (approche néopositiviste). Quand bien même il semblerait que le processus méthodologique puisse être plus ou moins objectif

selon le paradigme retenu, on évoquera à titre de clin d'œil la pensée nietzschéenne, faisant tomber du même coup ce dualisme sujet-objet: *Il n'y a pas de faits, il n'y a que des interprétations*. Par conséquent, au-delà de ces paradigmes qui risquent de scléroser les chercheurs dans un débat épistémologique, les méthodes qualitatives, quantitatives et mixtes peuvent s'avérer d'une utilité et d'une pertinence mutuelles, notamment dans le cadre de la recherche appliquée. Ces méthodes permettent donc avant tout de soutenir le processus de réflexion chez le chercheur ou l'équipe de recherche en fonction du contexte dans lequel s'inscrit le projet de recherche, soit la question de recherche, le phénomène et la population à l'étude.

Dans un domaine appliqué comme la réadaptation au travail, la psychologie du travail ou d'autres disciplines en sciences sociales, humaines et de la santé, il est parfois complexe de déterminer clairement l'approche méthodologique requise pour répondre aux questions de recherche émises par le chercheur ou son équipe. D'ailleurs, les réflexions sous-jacentes à ces questions de recherche peuvent parfois laisser le chercheur perplexe, obligé de faire des choix précis, par exemple quant à la méthode la plus pertinente à utiliser dans tel projet de recherche. La méthode est probablement la pierre angulaire de la réalisation effective du projet, car c'est elle qui facilite l'accès à la construction de la connaissance ou encore à la vérification d'hypothèses. La méthode est donc bien ce chemin (du grec *hodos*) que parcourt le chercheur, chemin méthodologique et épistémologique qui est parfois ardu pour ce dernier. Ces difficultés peuvent être liées au fait que certains départements d'université visent parfois exclusivement l'enseignement d'un paradigme (p. ex., positiviste) et d'une approche méthodologique (p. ex., quantitative) au détriment d'une autre approche (p. ex., constructiviste/qualitative). Aussi, de nombreux ouvrages orientent la présentation des méthodes en divisant systématiquement celles de nature qualitative de celles qui sont quantitatives, renforçant ainsi toujours un peu plus cette exclusion mutuelle des chercheurs, ceux qui sont dits qualitatifs de ceux qui sont dits quantitatifs ou, pour l'énoncer autrement, ceux qui sont dits constructivistes de ceux qui sont dits positivistes ou néopositivistes. Bien heureusement, et toutes choses étant égales par ailleurs, ces clivages ont tendance à s'estomper, voire à disparaître, car les méthodes qualitatives sont couramment utilisées de façon conjointe avec les méthodes quantitatives et vice versa. D'ailleurs, au-delà d'un projet de recherche particulier bien circonscrit, comment penser l'avancement des connaissances pour une discipline sans l'usage des méthodes qualitatives et quantitatives? Ne sont-elles pas complémentaires? Prenons l'exemple du développement d'une théorie, qui devra ensuite être testée et sur laquelle il faudra encore revenir pour la peaufiner et ainsi toujours mieux l'appréhender. Même s'il est parfois préférable d'utiliser une approche quantitative ou une approche

qualitative dans certains projets de recherche bien circonscrits, les méthodes qualitatives et quantitatives, ensemble, sont sans conteste requises pour développer nos connaissances dans la recherche appliquée.

Dans notre ouvrage collectif, où le titre renvoie aux méthodes qualitatives, quantitatives et mixtes, avec la collaboration de nombreux auteurs ayant une expertise dans l'utilisation de certaines d'entre elles, nous souhaitons offrir un large éventail de ces méthodes, illustrées par un devis de recherche particulier ou une analyse spécifique. Les directeurs de cet ouvrage se lancent aussi le défi d'établir des ponts entre les divers types de méthodes et de les rendre plus accessibles au lecteur, non seulement en proposant de les concevoir comme complémentaires et non mutuellement exclusives (p. ex., un chapitre renvoie à d'autres analyses exposées dans d'autres chapitres), mais aussi en offrant des applications détaillées et concrètes relatives aux champs d'intérêt de recherche des auteurs de ce livre, soit la santé psychologique/mentale, la santé physique et la réadaptation pour n'en nommer que quelques-unes. Ces liens méthodologiques et applications concrètes permettront au lecteur non seulement de mieux saisir une approche méthodologique en la comparant avec d'autres, mais aussi de reproduire le cheminement méthodologique emprunté par les auteurs. Autrement dit, l'objectif de ce livre n'est pas de donner une procédure de choix systématique ou une recette – à savoir l'analyse qu'il faut utiliser pour répondre à telle question de recherche en particulier –, mais plutôt d'éclairer le chercheur sur les différents chemins possibles pour la conduite du processus de recherche, le tout appuyé par des illustrations aisées à reproduire.

Dans le cadre de ce livre, nous avons choisi de commencer par les méthodes qualitatives, car elles nous semblent à la base même du développement d'un concept ou d'une théorie pour ensuite décrire des méthodes quantitatives. En ce sens, les méthodes qualitatives : recherche descriptive interprétative, phénoménologie, approche ethnographique, étude de cas et théorisation ancrée feront l'objet des premiers chapitres de ce livre. De plus, les chapitres sont présentés dans un ordre qui reflète aussi leurs similarités ou complémentarités. Par exemple, l'analyse de concept, la revue de littérature systématique et la métaanalyse sont des méthodes qui sont complémentaires, comme peuvent l'être, dans le domaine de l'évaluation de programme, l'analyse d'implantation des programmes et l'essai contrôlé randomisé pour l'analyse des effets, appuyées éventuellement par des analyses économiques. Pour ce qui est des similarités, nous avons regroupé dans cet ouvrage collectif les méthodes qui renvoient à l'utilisation de la notion de groupes d'individus, présentées dans les chapitres qui portent sur le groupe de discussion focalisée, l'approche Delphi, la technique TRIAGE,

l'analyse de classification par regroupement, les analyses de variance et l'analyse discriminante, qui toutes ont bien évidemment leurs caractéristiques propres, mais dont le groupe ou le regroupement d'individus est probablement le point de convergence. D'autres chapitres portent sur des méthodes qui peuvent impliquer la notion du temps ou de niveaux hiérarchiques comme les analyses de régression (linéaire et logistique), les analyses de survie, les analyses multiniveaux ou la modélisation par équations structurelles. Le lecteur aura aussi le loisir de faire un saut quantique, mais possible, entre la théorisation ancrée, la phénoménologie et la modélisation par équations structurelles ou les analyses factorielles dont l'objectif est de mieux articuler des liens conceptuels soit en les coconstruisant avec la population à l'étude, soit en les testant sous forme d'hypothèses. Enfin, des chapitres de nature à la fois qualitative et quantitative seront aussi présentés, comme la cartographie de concepts, la conception et la validation d'outils de mesure, les méthodes mixtes et les approches de recherche participatives.

Bonne découverte !

CHAPITRE 1

LA RECHERCHE DESCRIPTIVE INTERPRÉTATIVE

Description des besoins psychosociaux de femmes à la suite d'un résultat anormal à la mammographie de dépistage du cancer du sein

Frances Gallagher

FORCES

- Son utilité réside dans la production de descriptions des phénomènes rencontrés dans la pratique clinique.
- Elle génère une description en contexte des phénomènes à l'étude.
- Son caractère interprétatif permet de comprendre la signification des phénomènes selon le point de vue des personnes concernées par ces phénomènes.

LIMITES

- À ce jour, elle n'est pas balisée de façon consensuelle et a fait l'objet de peu d'écrits.
- Mal utilisée, elle pourrait se limiter à une description réductrice des phénomènes, qui n'intégrerait pas la dimension interprétative ou subjective de ceux-ci.
- Elle n'est pas indiquée pour le développement d'une théorie.

L'accès à une description des phénomènes est essentiel à la compréhension des situations rencontrées en clinique par les professionnels de la santé. À titre d'exemple, pensons aux besoins psychosociaux et au bien-être de la clientèle, au vécu des intervenants face à des situations cliniques bouleversantes et au point de vue des personnes concernant les soins, les thérapies ou les traitements offerts. Les descriptions en profondeur des phénomènes, tels qu'ils se manifestent en milieu naturel, et construites à partir du point de vue des personnes concernées permettent de comprendre ces phénomènes et d'en saisir les composantes et les variations (Corbin et Strauss, 2008; Loiselle, Profetto-McGrath, Polit et Beck, 2007; Thorne, 2008).

En pareils cas, la recherche qualitative descriptive interprétative est indiquée pour le développement de connaissances pertinentes pour les disciplines professionnelles intéressées par la santé mentale comme le sont la psychologie, les sciences infirmières, la médecine, le service social, l'ergothérapie et bien d'autres. Toutefois, très peu d'écrits portent sur ce dispositif de recherche, bien qu'il soit très utile au développement des connaissances. Le contenu de ce chapitre est notamment inspiré des écrits de Thorne (2008), de Thorne, Reimer Kirkham et O'Flynn-Magee (2004) et de Sandelowski (2000, 2010), des auteures ayant décrit et discuté de ce type de dispositif.

Le premier volet de ce chapitre présente dans l'ordre une définition de ce qu'est la recherche qualitative descriptive et ses fondements, les indications, les méthodes d'échantillonnage, de collecte et d'analyse des données, le type de résultats qui en découlent, et son utilité. Le second volet propose une illustration appliquée de ce dispositif de recherche.

1. CE QU'EST LA RECHERCHE DESCRIPTIVE INTERPRÉTATIVE

La recherche descriptive interprétative est un dispositif important en recherche qualitative (Sandelowski, 2000, 2010; Thorne, 2008). Cette forme de recherche sert à dépeindre un phénomène, ses propriétés, ses composantes et ses variations, à l'expliquer et à rendre compte de sa signification (Thorne, 2008). Il s'agit d'une démarche inductive de recherche attentive à la complexité des phénomènes humains et qui met en valeur la subjectivité humaine. La nature interprétative de la description renvoie à la quête de sens, de relations entre les composantes et d'agencements ou de configurations (*patterns*) (Thorne, 2008). À titre d'exemple, il peut s'agir de dégager des liens entre la qualité du milieu de vie des personnes et de leur accessibilité à des ressources en santé mentale pour mieux comprendre leurs besoins.

Qualifiée d'éclectique (Thorne, 2008), la recherche descriptive interprétative puise dans les diverses méthodes qualitatives pour explorer en contexte des sujets d'intérêt pour la clinique. Les résultats d'une telle recherche enrichissent le corpus de connaissances disciplinaires et soutiennent une pratique fondée sur des résultats probants (Sandelowski, 2008). Elle ne constitue pas une alternative plus facile aux méthodes classiques de recherche qualitative (Sandelowski, 2008, 2010; Thorne, 2008); elle donne plutôt un nom à de l'excellente recherche (Thorne, 2008).

Thorne (2008) et Sandelowski (2008, 2010) précisent que la recherche descriptive interprétative se veut davantage un cadre méthodologique que la «prescription» d'un processus de recherche ou un devis de recherche au sens classique du terme. Toutefois, Thorne (2008) présente tous les éléments du plan d'ensemble permettant de structurer une étude ainsi que les moyens d'en contrôler la qualité, ce qui, somme toute, correspond à un devis de recherche (Fortin, 2010; Loiselle *et al.*, 2007). À preuve, le cadre repose sur les fondements de la recherche qualitative et oriente les décisions méthodologiques relatives à la position du chercheur, à l'échantillonnage, à la collecte et à l'analyse des données ainsi qu'aux mécanismes préconisés pour en assurer la rigueur scientifique, de façon à répondre à des questions de recherche et à produire des résultats significatifs pour les disciplines appliquées.

1.1. Fondements de la recherche descriptive : de nouvelles lunettes

La recherche descriptive interprétative repose sur les fondements de la recherche qualitative (Creswell, 2013; Sandelowski, 2000, 2010; Thorne, 2008), lesquels sont présentés dans les lignes qui suivent. Elle représente la seule façon d'obtenir une description interprétative détaillée et holiste du phénomène dans son contexte naturel (Creswell, 2013; Denzin et Lincoln, 2011; Laperrière, 1997). Elle mise sur la prise en compte de la subjectivité humaine en portant attention au point de vue du chercheur et à celui des participants (Creswell, 2013; Laperrière, 1997). La recherche descriptive interprétative accorde une importance à la découverte, à l'exploration et à la quête de la signification des phénomènes selon la perspective des participants. Le dispositif de recherche est souple, évolutif et itératif (Creswell, 2013; Pelaccia et Paillé, 2010).

1.2. Problème et question de recherche

La description interprétative répond au besoin de comprendre la signification des phénomènes. La recherche descriptive interprétative s'intéresse aux phénomènes humains dont l'état actuel des connaissances ne permet pas une compréhension en contexte et en profondeur (Thorne, 2008). Elle répond à des préoccupations de nature clinique et disciplinaire qui nécessitent une description émergeant d'un processus inductif. Il va sans dire que les cliniciens ont besoin de réponses à des problèmes ou à des préoccupations qu'ils rencontrent en clinique concernant, entre autres, ce que les personnes utilisatrices de services pensent, ce qu'elles ressentent, leur réalité, leur expérience de santé et de maladie au quotidien, leurs interactions avec les professionnels de la santé, ainsi que la propre expérience des cliniciens en tant que soignants ou thérapeutes (Miller et Crabtree, 2003).

La recherche descriptive interprétative permet de répondre à des questions ou à cerner des objectifs visant à dépeindre un phénomène, à définir ses composantes et à l'expliquer. Ces questions correspondent à des préoccupations disciplinaires (Thorne, 2008), émergeant dans bien des cas de la clinique et utiles pour la pratique professionnelle.

Voici quelques exemples de questions inspirées de Sandelowski (2000) permettant de découvrir les éléments composant un phénomène : Quelles sont les préoccupations des personnes au sujet de… ? Quels sont leurs besoins relatifs à… ? Quelles sont leurs réactions (pensées, sentiments, actions) par rapport à un soin, à une approche thérapeutique, à un événement… ? Pour quelles raisons les personnes utilisent-elles ou n'utilisent-elles pas un service ? Quels facteurs facilitent et contraignent le rétablissement à la suite de… ?

La recherche descriptive ne se limitant pas à la définition des composantes d'un phénomène donné (Sandelowski, 2010 ; Thorne, 2008), des questions de nature interprétative s'ajoutent nécessairement aux questions plus descriptives. En voici quelques exemples : Comment ces éléments sont-ils reliés les uns aux autres ? Comment ces éléments agissent-ils ? Quelle configuration peut-on dégager à partir de ces thèmes ?

Ce type de questions de recherche, pourtant toutes fort pertinentes pour les disciplines professionnelles ou appliquées, ne correspond pas aux visées des approches plus traditionnelles de recherche qualitative. À titre d'exemple, ce ne sont pas des questions centrées directement sur l'essence d'un phénomène (voir le chapitre 2 de cet ouvrage), la culture d'un groupe (voir le chapitre 3 de cet ouvrage), la description d'un système bien délimité (voir le chapitre 4 de cet ouvrage), ou le développement d'une théorie ou la modélisation d'un processus social (voir le chapitre 5 de cet ouvrage).

À première vue, on peut toutefois confondre la recherche descriptive interprétative avec d'autres types de recherche qualitative. Pour illustrer la distinction entre ce type de recherche et la phénoménologie descriptive, prenons l'exemple des besoins psychosociaux des femmes ayant reçu un résultat anormal à la mammographie de dépistage du cancer du sein (voir l'illustration en deuxième partie du présent chapitre). La phénoménologie descriptive ou transcendantale donnerait lieu à une compréhension profonde de la structure du phénomène étudié. Ainsi, les résultats d'une phénoménologie descriptive, par la réduction eidétique (c.-à-d. l'exercice d'aller au-delà des faits particuliers ou individuels), nous renseigneraient sur l'essence pure des besoins, c'est-à-dire ce qui constitue le caractère universel de ceux-ci. Comme illustré dans la deuxième partie du présent chapitre, la recherche descriptive interprétative permettrait de cerner les types de besoins ressentis, leurs caractéristiques, dans quel contexte ils se font sentir, lesquels sont comblés, lesquels ne le sont pas. De plus, elle apporterait un éclairage sur le vécu émotif des femmes associé à cette expérience, pour une compréhension holiste de leurs besoins.

La recherche descriptive interprétative partage des caractéristiques communes avec l'étude de cas, dont la quête d'une description en contexte, en profondeur et holiste ainsi que l'à-propos de recourir à diverses méthodes de collecte et d'analyse des données. La recherche descriptive interprétative se distingue de l'étude de cas par le fait qu'elle sert à dépeindre des phénomènes humains (p. ex., ressentis, besoins, vécu face à une situation), alors que l'étude de cas vise plutôt à décrire de façon détaillée et complète non pas un problème ou un thème, mais une entité (cas) telle qu'une ou plusieurs personnes, un événement, un programme ou une organisation (Stake, 1995). Aussi, l'étude de cas nécessite habituellement plusieurs méthodes de collecte des données pour obtenir une description exhaustive du ou des cas étudiés. Ainsi, dans une étude de cas inspirée de notre exemple, nous pourrions étudier le cas des femmes (cas étant des personnes) ayant obtenu un résultat anormal à leur mammographie de dépistage du cancer du sein au lieu de nous pencher plus précisément sur leurs besoins psychosociaux (problème ou thème).

1.3. Position du chercheur

Le processus de la recherche descriptive interprétative implique la divulgation et la prise en compte du positionnement particulier du chercheur (Thorne, 2008), ce dernier étant en quelque sorte un instrument de recherche. Il importe donc que ses positions théoriques et disciplinaires soient clairement indiquées (Thorne, 2008), et qu'une analyse constante

de son implication sociale et émotive dans la relation avec les personnes participant à la recherche soit effectuée (Laperrière, 1997). Cette précaution permet la distanciation nécessaire pour garantir une description interprétative du phénomène qui sera considérée comme une contribution empirique significative et crédible. Ceci implique qu'il y ait « concordance entre les données empiriques et leur interprétation » (Laperrière, 1997, p. 377) et que le point de vue des participants soit bien rendu. Notons que la tenue d'un journal de bord réflexif, les notes méthodologiques, les discussions en équipe, la coanalyse des données sont tous des exemples de moyens pour considérer la position du chercheur.

1.4. Échantillon

L'échantillon se compose de personnes « expertes » du phénomène étudié (p. ex., personnes atteintes de schizophrénie parce qu'elles connaissent cette réalité), d'événements (p. ex., séances de thérapie de groupe), de documents (p. ex., écrits, vidéos, photos), de lieux et de moments (p. ex., sur une unité de soins à divers moments de la journée), toutes choisies parce qu'elles contribueront à la compréhension et à une description significative des diverses perspectives du phénomène (Sandelowski, 1995 ; Thorne, 2008). En outre, la problématique et les questions ou objectifs de la recherche guident les décisions concernant l'échantillon. Un échantillon de convenance peut constituer une excellente source d'information (Thorne, 2008). Cependant, une stratégie d'échantillonnage intentionnel (Sandelowski, 1995), à variation maximale, a l'avantage de déterminer à l'avance des caractéristiques (p. ex., sociodémographiques), des conditions (p. ex., années d'expérience en santé mentale, première rencontre avec l'intervenant) ou des contextes (p. ex., moments précis de la trajectoire de la maladie, lieux d'intervention) à considérer afin de couvrir plusieurs aspects du phénomène (Thorne, 2008).

La taille de l'échantillon dans le cadre d'une recherche descriptive interprétative est variable, aucun nombre précis de sources d'information n'étant recommandé comme tel (Thorne, 2008). Elle doit néanmoins favoriser l'élaboration d'une description suffisamment complète, réaliste et contextualisée pour permettre de bien cerner le phénomène étudié à partir des résultats présentés. La taille exacte peut difficilement être déterminée à l'avance. Cependant, le critère de saturation empirique (Pires, 2007), aussi appelée saturation des données, peut orienter le chercheur à ce sujet. La saturation empirique désigne les éléments sur lesquels se base le chercheur pour juger que les dernières données recueillies n'apportent pas d'information nouvelle ou différente pour justifier la poursuite de la collecte des données auprès de nouvelles sources d'information. La notion de saturation

doit toutefois être utilisée judicieusement (Thorne, 2008), et le chercheur est invité à expliquer comment il a établi le niveau de saturation atteint (O'Reilly et Parker, 2013). Certains facteurs peuvent cependant influer sur la taille de l'échantillon, notamment l'accès à la population, l'homogénéité de l'échantillon, la rareté du phénomène, les ressources et le temps à la disposition du chercheur. Le nombre de participants dans les études descriptives interprétatives ayant fait l'objet de publications varie. À titre d'exemple, il va de 11-12 participants (Thorne, Con, McGinness *et al.*, 2004; Thorne, McGinness, McPherson *et al.*, 2004) à 21-22 (Maheu et Thorne, 2008; Lasiuk, Comeau et Newburn-Cook, 2013) et même jusqu'à 60 participants (Thorne *et al.*, 2010).

1.5. Collecte et analyse des données

La collecte des données visant à décrire la signification d'un phénomène pertinent pour la clinique et la discipline (p. ex., besoins, opinions, réactions à des interventions) sert notamment à explorer la perspective des participants et à découvrir les multiples réalités en contexte eu égard au phénomène étudié (Thorne, 2008). Le chercheur peut puiser parmi la gamme de méthodes de collecte des données habituellement utilisées en recherche qualitative: l'entrevue individuelle semi-dirigée et l'entrevue informelle, l'observation (participante ou non), l'entrevue de groupe, l'examen de documents et de diverses productions humaines (p. ex., vidéos, photos). Nous les présenterons brièvement dans les lignes qui suivent et invitons le chercheur à consulter plusieurs excellentes références disponibles sur ces méthodes de collecte des données (entre autres, Kvale et Brinkmann, 2009; Poupart, 1997).

L'*entrevue de recherche* se définit comme une méthode de conversation professionnelle à l'intérieur de laquelle le savoir est construit dans le cadre d'une interaction entre la personne interviewée et l'intervieweur (Kvale et Brinkmann, 2009). L'entrevue semi-dirigée est animée de façon souple par l'intervieweur, qui pose des questions ouvertes visant à encourager la personne interviewée à décrire sa réalité sociale, ce qu'elle pense, ce qu'elle vit ou ce qu'elle a vécu tout en portant une attention à l'interprétation de la signification du phénomène étudié (Kvale et Brinkmann, 2009; Poupart, 1997). L'intervieweur ne se limite pas à poser des questions; il doit également découvrir quelles questions poser et quand les poser (Paillé, 1991). Plus précisément, l'intervieweur joue le rôle d'un facilitateur pour encourager la personne à verbaliser, à s'exprimer. Il doit posséder des compétences affectives (p. ex., écoute active, respect, sensibilité), professionnelles (p. ex., capacité à guider l'entrevue, à poser les questions pertinentes, à faire des

synthèses et des transitions) et techniques (p. ex., habiletés de communication telles que la reformulation) (Savoie-Zajc, 2009). L'entrevue semi-dirigée est réalisée à l'aide d'un guide ou d'un schéma comportant des questions ouvertes, auxquelles s'ajoutent des questions d'approfondissement en cours d'entretien. Elle se déroule à un moment et dans un lieu à la convenance de la personne interviewée, et elle est enregistrée en mode audionumérique pour assurer l'intégrité des données recueillies. L'intervieweur doit faire preuve de réflexivité pour diminuer son influence et se centrer sur la réalité de la personne interviewée.

L'*observation* de diverses activités humaines ou d'événements (p. ex., rencontres entre intervenant et client, pratiques professionnelles des intervenants d'une ressource alternative) constitue une méthode de collecte de données utile dans une recherche descriptive interprétative. Elle permet d'enrichir la description en documentant, par exemple, le contexte et les actions des personnes (p. ex., intervenants, clients, proches aidants). Combinée à l'entrevue qualitative, elle permet de saisir ce qui sous-tend les actions observées, de comprendre le sens de ces actions selon le point de vue des participants (Jaccoud et Mayer, 1997 ; Thorne, 2008). L'observateur peut adopter une position externe ou non participante aux situations observées, ou s'intégrer à divers degrés aux activités afin de saisir de l'intérieur le sens des actions. Il importe de choisir les sites et situations avec minutie afin de cibler les points à observer de façon à éviter l'éparpillement (Jaccoud et Mayer, 1997). Le chercheur consigne ses observations, ou notes de terrain, dans une grille conçue à cette fin ou dans un style libre dans un journal. Ses notes de terrain renseignent sur le site, le contexte temporel, les personnes présentes, les actions et les interactions. L'observation nécessite du temps sur le terrain ainsi qu'une relation neutre mais suffisamment cordiale pour favoriser la confiance des participants. L'observateur doit documenter, dans son journal réflexif, l'effet de sa présence sur le terrain et l'effet du terrain sur son travail de terrain. Le chercheur qui réalise des observations dans un contexte qui lui est familier doit toutefois être très attentif aux détails (Thorne, 2008) qu'il ne remarque plus afin de fournir une description la plus juste et complète possible.

L'*entrevue de groupe*, aussi appelée groupe de discussion (Geoffrion, 2009) ou *focus group* (Touré, 2010), se définit comme une « technique d'entrevue qui réunit de six à douze participants et un animateur, dans le cadre d'une discussion structurée, sur un sujet particulier » (Geoffrion, 2009, p. 391). Il s'agit d'une activité collective, qui mise sur l'interaction au sein du groupe pour documenter la motivation, des opinions, des attitudes, des recommandations ou de nouvelles idées par rapport au phénomène étudié. De surcroît, elle donne l'occasion de dépeindre ces interactions et

de capter le langage utilisé pour parler d'un phénomène (Kitzinger, 1994; Touré, 2010). L'entrevue de groupe, habituellement enregistrée en version audionumérique, est réalisée à l'aide d'un guide ou d'un canevas. Un observateur peut assister l'animateur pour raffiner les questions d'approfondissement et prendre des notes de terrain (p. ex., langage non verbal) (voir le chapitre 12 de cet ouvrage).

Des *documents* choisis pour leur capacité à documenter des aspects du phénomène et ainsi contribuer à leur description peuvent constituer des sources d'information pertinentes. Il peut s'agir de documents écrits (p. ex., guides de pratique, procédures de soins), d'enregistrements audio ou vidéo, de photos ou d'illustrations. L'examen du contenu de sites Internet (p. ex., sites consultés par les personnes utilisatrices de services) et de forums de discussion représente une autre avenue à explorer. Il importe de déterminer la nature des données à saisir et la modalité de la collecte des données afin d'assurer la qualité de la saisie.

La triangulation des sources ou des méthodes de collecte des données favorise l'obtention d'une description holiste et significative. Le choix des méthodes dépend du temps et des ressources disponibles. En terminant, notons qu'il importe de bien connaître les forces et les limites de chacune de ces méthodes.

1.6. Traitement et analyse des données

L'analyse qualitative des données est évolutive et réalisée tout au long du processus de recherche. De cette façon, le chercheur saisit en contexte et progressivement le sens des données recueillies. Fort de la compréhension émergeant de ses premières analyses, il est en mesure de prendre des décisions éclairées pour orienter la stratégie d'échantillonnage et mieux centrer la collecte des données. D'une part, une méthode d'analyse appropriée doit faciliter le repérage des composantes du phénomène étudié, le regroupement de ces composantes et la mise en évidence des relations qui existent entre elles. D'autre part, elle vise une compréhension de la signification du phénomène. Plusieurs méthodes d'analyse qualitative des données répondent aux visées d'une recherche descriptive interprétative. À titre d'exemple, citons les travaux de Miles et Huberman (2003), de Miles, Huberman et Saldaña (2014), de Crabtree et Miller (1999) et de Paillé et Mucchielli (2010). Thorne (2008) suggère, entre autres, la méthode d'analyse développée en théorisation ancrée pour l'accent mis sur la découverte des relations entre les catégories. Les méthodes d'analyse qualitative comportent des similitudes et elles ne se limitent pas à un exercice de classification. Il est d'abord

recommandé de doubler et de sauvegarder le matériel et de transcrire rigoureusement l'ensemble des données recueillies au cours des entrevues et des observations afin d'en faciliter l'analyse.

Le processus proposé par Miles et Huberman (2003) et Miles *et al.* (2014) pour l'analyse des données est décrit dans le présent chapitre, puisque cette méthode est largement utilisée dans les recherches du domaine de la santé. Selon ces auteurs, le processus d'analyse comporte essentiellement trois flux concourants d'activités : 1) la condensation des données ; 2) la présentation des données, qui implique, entre autres, la mise en relation des thèmes ; 3) l'élaboration de conclusions et la vérification de celles-ci. Des références à d'autres auteurs sont intégrées pour compléter la description.

1.6.1. Condensation des données

En vue de la condensation des données, le processus d'analyse s'amorce par une lecture et une relecture de la transcription des données et l'écoute de l'enregistrement des entrevues afin de s'imprégner des données et de développer une vision globale du point de vue des participants. Le chercheur peut immédiatement rédiger une fiche-synthèse relatant les éléments qui ont retenu son attention de même que des questions ou problèmes susceptibles d'orienter la suite du travail sur le terrain.

Le premier cycle de la condensation des données (Miles *et al.*, 2014) consiste à examiner les données pour y repérer les extraits significatifs et à dégager les thèmes (appelés codes ou catégories selon les auteurs) qui y sont abordés. Plus précisément, un thème consiste en une courte expression qui capte « l'essentiel d'un propos ou d'un document » (Paillé et Mucchielli, 2010, p. 164). Le thème doit rendre compte du sens des propos, d'où la pertinence d'écouter l'entrevue pour saisir des indices (p. ex., intonation, silences, débit verbal) permettant d'avoir une vue plus juste du témoignage.

À cette fin, le chercheur utilise une grille d'analyse composée de thèmes qui serviront à décrire et à expliquer le phénomène à l'étude. Cette grille peut être construite au fur et à mesure qu'il découvre des thèmes à la lecture du matériel recueilli. En s'inspirant de ses objectifs de recherche, des guides d'entrevues et de la recension des écrits, le chercheur peut aussi élaborer une grille initiale d'analyse constituée de thèmes prédéterminés auxquels des thèmes révélés en cours d'analyse s'ajouteront. Une définition des thèmes illustrée par des exemples tirés du compte rendu intégral (*verbatim*) ou des observations accompagne la grille d'analyse.

Le deuxième cycle de la phase de condensation (Miles *et al.*, 2014) implique que le chercheur cerne les liens entre les thèmes, construisant progressivement un arbre thématique en regroupant les thèmes sous des thèmes plus généraux, à caractère explicatif ou de niveau conceptuel plus élevé. Cet arbre thématique sous la forme d'une structure unifiée (Miles *et al.*, 2014) représente une ébauche des éléments qui serviront à la description du phénomène. Aussi, tout au long du processus d'analyse, le chercheur rédige des mémos servant à documenter ses réflexions analytiques (p. ex., hypothèse de l'existence d'un lien entre deux thèmes ou codes) ainsi que ses décisions méthodologiques (p. ex., recruter un type de participant pour mieux définir certains codes).

1.6.2. Présentation des données

La présentation des données consiste à utiliser diverses stratégies pour mettre en évidence les relations entre les thèmes, dans l'optique d'aller au-delà de la description, pour une compréhension plus approfondie, plus interprétative du phénomène. Elle est progressive et gagne à être amorcée de façon concomitante aux activités de condensation des données. Il s'agit de présenter « l'information sous une forme ramassée et ordonnée » (Miles et Huberman, 2003, p. 174), ce qui permet au chercheur de poser un regard plus englobant sur les données et les résultats obtenus (c.-à-d. sur les thèmes ou catégories) grâce à des activités de condensation des données (c.-à-d. les données et thèmes concernant une entrevue ou une observation et sur l'ensemble du matériel).

Miles *et al.* (2014) proposent la présentation des données sous forme de matrices (c.-à-d. de tableaux), de schémas ou de réseaux. Le format de présentation choisi doit contribuer à répondre à la question de recherche. C'est un exercice créatif, qui doit être réalisé avec rigueur, et qui peut nécessiter plusieurs itérations avant de trouver le format qui représente le mieux les relations et l'interprétation juste du phénomène.

À titre d'exemple, soulignons l'utilité des matrices ou schémas contextuels (p. ex., permet de visualiser dans quelles circonstances la personne éprouve davantage d'anxiété – lors d'interactions avec les proches ou autres), chronologiques (p. ex., apporte une vision temporelle des épisodes d'anxiété), à regroupement conceptuel (p. ex., donne un aperçu des liens entre les catégories définissant les stratégies d'adaptation utilisées et celles relatives à l'estime de soi), et explicatifs. Miles et Huberman (2003) associent les matrices ou schémas explicatifs à la construction du sens. Cette présentation des données implique la construction d'un scénario, d'une configuration, d'une représentation du phénomène qui fait sens. À titre

d'exemple, il pourrait s'agir d'un tableau présentant les émotions dominantes, la façon dont les personnes expriment ces émotions, les circonstances où elles sont ressenties, les réactions et le regard que chacun des participants porte sur son expérience de dépression. Des caractéristiques sociodémographiques pourraient aussi être ajoutées à cette représentation. En examinant le contenu du tableau, le chercheur peut mieux comprendre la complexité du phénomène qu'il veut décrire. Il pourrait, par exemple, observer que certains participants anticipent les interactions avec les membres de l'équipe de santé mentale surtout lorsqu'il est question d'un sujet en particulier et à propos duquel ils se sentent démunis. Ce genre de constat permet de construire graduellement une description interprétative du vécu des participants.

Le processus de recherche étant itératif, la présentation des résultats effectuée de façon progressive a l'avantage de nourrir la collecte des données ainsi que les autres activités que sont la condensation des données, et l'élaboration et la vérification des conclusions.

1.6.3. Élaboration et vérification des conclusions

Le troisième courant d'activités d'analyse implique l'élaboration et la vérification des conclusions qui émergent des deux autres courants d'activités. Dès le début de la collecte des données, le chercheur développe une compréhension du sens du phénomène étudié. Entre autres, il note des régularités (p. ex., anticipation de scénarios sombres chez tous les participants), des particularités (p. ex., les plus jeunes participants expriment de la frustration face à un type de situation) et des explications (p. ex., l'information concernant la médication manque de clarté, ce qui entraîne de la méfiance, le sentiment d'être manipulé et le refus de la prendre). Des conclusions émergent d'abord de l'analyse des données recueillies lors d'une entrevue, d'une observation, d'un *focus group* ou d'une autre source d'information et ainsi de suite jusqu'à l'élaboration de conclusions découlant d'une compréhension de l'ensemble du matériel analysé.

Ces conclusions doivent faire l'objet de vérifications au fur et à mesure, autant les premières impressions après la rédaction d'une fiche synthèse que les conclusions à un stade plus avancé du processus d'analyse. Plusieurs moyens peuvent être utilisés : une distanciation par la rédaction de son journal réflexif, les discussions régulières entre les membres de l'équipe de recherche et la coanalyse des données pour un échange constructif des idées et une meilleure compréhension du sens de l'objet d'étude, le retour auprès

des participants pour valider la compréhension de leur point de vue, ainsi que le retour constant aux données brutes pour assurer une concordance entre les données et les conclusions.

Dans le cadre d'une analyse qualitative, seule une activité humaine permet de s'approprier et de comprendre le point de vue des participants à la recherche. Cependant, les logiciels soutenant l'analyse qualitative (p. ex., NVivo) sont très utiles pour la gestion des données. Ils sont particulièrement pratiques pour générer une variété de rapports d'analyse. Parmi ces possibilités, mentionnons les rapports d'analyse concernant un code et ses unités de sens (p. ex., tous les extraits sous le code « médicament synonyme de poison »), l'ensemble des codes et unités de sens pour une entrevue ou observation ou document, ou un schéma des liens entre les codes. L'utilisation d'un tel programme facilite la vérification des conclusions, entre autres par la possibilité de générer des rapports par sous-groupes de participants (p. ex., pour vérifier votre conclusion que les hommes semblent avoir une vision plus positive de la médication) et des rapports pour chacun des analystes (cocodification).

L'analyse qualitative des données est un processus itératif, effectuée en concomitance avec la collecte des données et même avec la constitution de l'échantillon. En effet, les conclusions de l'analyse peuvent appuyer, par exemple, la pertinence d'explorer le point de vue de personnes dont l'expérience serait différente de celle des participants déjà recrutés pour une meilleure description du phénomène (p. ex., avoir plus d'hommes dans un échantillon).

2. ILLUSTRATION DE LA RECHERCHE DESCRIPTIVE INTERPRÉTATIVE

Cette partie se veut une illustration de l'utilisation de la recherche descriptive interprétative dans le cadre d'une étude (Doré *et al.*, 2013) portant sur les besoins psychosociaux de femmes en attente d'examens complémentaires à la suite de leur mammographie de dépistage qui présentait un résultat anormal. La présentation de la recherche dépeint le contexte et les objectifs de la recherche, la position des chercheuses, le bien-fondé de l'approche de la description interprétative, les méthodes de collecte et d'analyse des données utilisées, ainsi que quelques résultats de cette recherche et leur interprétation.

2.1. Contexte de l'étude

La prévalence du cancer du sein constitue une préoccupation majeure de santé dans plusieurs pays occidentaux. Il s'agit du principal type de cancer rencontré chez les femmes, affectant une femme sur huit au Canada (Statistique Canada, 2011). Face à la prévalence élevée de ce cancer chez les femmes, le ministère de la Santé et des Services sociaux du Québec a mis sur pied en 1998 le Programme québécois de dépistage du cancer du sein (PQDCS) afin de le déceler à un stade précoce et de diminuer la mortalité des femmes qui en sont atteintes. Ce programme de dépistage comprend, entre autres, une mammographie de dépistage pour les femmes âgées de 50 à 69 ans. Le résultat de la mammographie est transmis au médecin de famille, qui doit assurer le suivi requis. Ce type de programme peut toutefois engendrer de l'anxiété et de l'appréhension chez les femmes, particulièrement lorsqu'elles reçoivent un résultat anormal à la suite de la mammographie de dépistage, ce qui implique des examens complémentaires afin de déterminer le diagnostic final. Or, pendant la phase prédiagnostique, le soutien psychosocial disponible ne répond pas aux attentes des femmes (Demir *et al.*, 2008 ; Liao *et al.*, 2007, 2008 ; McCaughan et McSorley, 2007) qui vivent de l'inquiétude et de l'incertitude liées à leur santé et au sens menaçant que revêt cette maladie (Drageset, Lindstrom et Underlid, 2010 ; Heckman *et al.*, 2004). Comme les besoins des femmes qui traversent cette période ont peu retenu l'attention des chercheurs jusqu'à maintenant et afin que le PQDCS puisse mieux soutenir les femmes concernées, nous nous sommes penchées sur l'expérience des femmes quant à leurs besoins psychosociaux.

Deux de nos questions de recherche étaient plus descriptives : Quel est le vécu des femmes ayant reçu un résultat de mammographie de dépistage anormal ? Quelle est la nature des besoins des femmes en matière de soutien psychosocial et informatif ? Et la dernière question se voulait davantage interprétative afin de comprendre la signification de ces besoins selon la réalité des participantes au PQDCS : Comment peut-on expliquer ces besoins ?

2.2. Position des chercheuses

En tant que femmes, les chercheuses de l'équipe se sentaient concernées par le phénomène à l'étude. D'ailleurs, quelques-unes d'entre elles faisaient partie de la population visée par le PQDCS et avaient donc eu à passer une mammographie de dépistage. Afin d'assurer la distanciation nécessaire à la prise en compte de la réalité des participantes à la recherche, les chercheuses ont réalisé des retours réflexifs avec l'assistante de recherche qui interviewait

les participantes, ont discuté ensemble de leur compréhension du contenu des entrevues et des notes dans le journal personnel des participantes, ont coanalysé le matériel et ont rédigé systématiquement les idées et décisions méthodologiques. À titre d'exemple, les discussions entre les membres de l'équipe de recherche permettaient la verbalisation des émotions suscitées par le contenu des entrevues pour ensuite se concentrer essentiellement sur le point de vue exprimé par les participantes.

2.3. Justification de l'approche de la description interprétative

La recherche descriptive interprétative s'avérait une approche indiquée pour explorer et comprendre les besoins des femmes en attente d'examens complémentaires à la mammographie de dépistage dont le résultat montrait une anomalie. Plus précisément, elle nous permettait de dépeindre les besoins psychosociaux et informatifs à partir du point de vue des femmes concernées (subjectivité humaine) en tenant compte de leur vécu pendant la trajectoire prédiagnostique (contexte). Ainsi, les résultats attendus ne se limitaient pas à une simple définition : nous voulions décrire les besoins et les expliquer de façon à ce que les résultats reflètent les multiples réalités des participantes à la recherche et le sens qu'elles donnent à leurs besoins. Par ailleurs, soulignons que toutes les étapes de la recherche ont été réalisées en partenariat avec le Centre de santé des femmes et les responsables du PQDCS de la région où s'est déroulée l'étude, ce qui a contribué à répondre aux préoccupations de ces acteurs dans le domaine de la santé des femmes.

2.4. Participantes à la recherche

L'échantillon devait être composé de femmes suivies dans le cadre du PQDCS dans une région du Québec. Pour participer à la recherche, les femmes devaient être âgées entre 50 et 69 ans, avoir reçu un résultat anormal à la suite d'une mammographie récente et parler français. Elles étaient considérées comme des « expertes » du phénomène étudié, puisqu'elles étaient au cœur de la phase prédiagnostique. Elles pouvaient alors nous faire part de ce qu'elles vivaient et de leurs besoins au moment même où elles traversaient cette période cruciale de leur vie.

Un échantillonnage intentionnel, à variation maximale, a été effectué, ce qui nous a permis de déterminer à l'avance des caractéristiques des participantes à considérer pour couvrir le plus largement possible ce qui caractérise leurs besoins. Ainsi, une attention a été portée à la diversité pour ce qui est du lieu de résidence (milieux rural et urbain), de l'étape du processus

prédiagnostique (en attente d'examens complémentaires ou de résultats) et de l'expérience antérieure de santé (p. ex., kystes au sein, personne de l'entourage atteinte du cancer du sein). Bien qu'il soit difficile de déterminer à l'avance la taille de l'échantillon, nous pensions qu'une vingtaine de participantes nous permettrait d'atteindre un niveau acceptable de saturation des données. D'ailleurs, les besoins exprimés par les trois ou quatre dernières participantes corroboraient ceux exprimés par les participantes précédentes.

Après avoir obtenu l'approbation du comité d'éthique de la recherche, nous avons amorcé le recrutement des femmes admissibles à l'étude. Les moyens suivants ont été utilisés pour inviter les femmes à participer à l'étude : une lettre remise par le personnel du programme de dépistage après la mammographie ou lors de la visite médicale annonçant une anomalie à la mammographie, ainsi que des affiches dans les salles d'attente des centres participant au programme (c.-à-d. des lieux où les mammographies, échographies ou autres examens sont effectués) et des cliniques médicales de la région. De plus, l'assistante de recherche s'est rendue dans chacune des salles d'attente des cliniques de la région pour informer les femmes de la tenue de l'étude. L'assistante de recherche communiquait avec les femmes intéressées à recevoir plus d'information et elle répondait à leurs questions. Ensuite, une rencontre était fixée pour une entrevue avec celles ayant accepté de participer à l'étude. Avant de procéder à l'entrevue, le formulaire de consentement était lu et signé.

2.5. Collecte des données

Quatre modalités de collecte des données ont été utilisées dans l'optique de documenter la perspective des participantes, à savoir l'entrevue individuelle semi-dirigée, le journal des participantes, le groupe de discussion focalisée (*focus group*) et le journal des chercheuses. Deux entrevues semi-dirigées ont permis d'explorer auprès de chacune des participantes l'expérience d'avoir récemment reçu un résultat anormal à la suite d'une mammographie et leurs besoins en lien avec cette expérience. L'entrevue se déroulait le plus possible sous la forme d'une conversation dirigée avec souplesse par l'assistante de recherche à l'aide d'un guide d'entrevue semi-structuré. Ce guide comportait des questions ouvertes, avec la possibilité d'ajouter des questions d'approfondissement afin d'encourager et d'aider la personne à décrire sa réalité, son point de vue. Voici quelques exemples de questions posées lors de la première entrevue : « Racontez-moi comment vous avez appris que votre mammographie était anormale », « Racontez-moi ce que vous avez vécu depuis », « Parlez-moi de ce qui a été aidant ou moins aidant pour vous depuis que vous êtes en attente d'examens complémentaires ou de résul-

tats », « Qu'est-ce qui fait que telle chose a été aidante (ou non aidante) ? » Des questions telles que « Pouvez-vous me donner un exemple de… ? Que voulez-vous dire par… ? Dans quelles circonstances… ? » permettaient de comprendre le sens des propos des participantes et de saisir des éléments contextuels associés au vécu et aux besoins des femmes. Cette première entrevue, réalisée dans un lieu au choix de la participante, d'une durée variant entre 45 et 90 minutes, a été enregistrée en version audionumérique. À la fin de cette entrevue, les participantes répondaient à un court questionnaire portant sur leurs caractéristiques sociodémographiques (p. ex., âge, milieu de vie, niveau de scolarité).

Quatre à six semaines plus tard, une deuxième entrevue, téléphonique cette fois, d'une durée variant entre 15 et 30 minutes, permettait de vérifier et de poursuivre notre compréhension du contenu de la première entrevue, de compléter ou de modifier certains éléments, et d'ajouter de nouvelles données. Soulignons qu'un résumé de la première entrevue était transmis par la poste à la participante, ce qui facilitait le déroulement de la deuxième entrevue. Après s'être assurée auprès de la participante que le moment pour l'entrevue téléphonique lui convenait, l'intervieweuse lui demandait comment elle allait, s'il y avait eu des changements depuis la première entrevue. Ensuite, elle lui demandait ce qu'elle pensait du résumé, si elle voulait y apporter des changements, des ajouts (besoins, ressentis, etc.) en lien avec son vécu plus récent. Elle profitait de l'occasion pour vérifier l'intérêt à participer au *focus group*. L'intervieweuse notait les propos des participantes sur un document prévu à cette fin.

Entre les deux entrevues, les participantes étaient invitées à mettre librement par écrit ce qu'elles vivaient dans un journal personnel (besoins, ce qui vous manque, ce qui vous aide, vos frustrations, vos peurs, vos difficultés) et à le retourner par la poste, dans une enveloppe préaffranchie, après la deuxième entrevue.

Une fois l'analyse du contenu des entrevues individuelles terminée (n = 2 entrevues × 20 participantes), une entrevue de type *focus group* a été réalisée auprès de sept participantes volontaires parmi celles ayant été interviewées individuellement. Cette entrevue de groupe, d'une durée de deux heures, enregistrée en version audionumérique, avait pour but de valider les résultats émergeant des entrevues individuelles et de favoriser les échanges entre les participantes sur leur expérience et leurs besoins. Après une mise en contexte, les participantes étaient invitées à partager les besoins qu'elles ressentaient pendant la période d'attente prédiagnostic, l'aide reçue et les ressources utilisées, et les suggestions d'aide ou de services pour les femmes qui ont à vivre une situation semblable.

Finalement, l'assistante de recherche tenait un journal dans lequel elle consignait ses notes de terrain, réflexions et impressions dans le but de compléter le contenu des entrevues et de documenter l'effet du chercheur et l'effet des participantes sur son travail en tant que membre de l'équipe de recherche.

2.6. Analyse des données

Une analyse qualitative de l'ensemble des données recueillies lors des entrevues individuelles en présence ($n = 20$), des entrevues téléphoniques ($n = 20$), des journaux des participantes ($n = 11$), du *focus group* a été effectuée à partir d'une approche inductive inspirée des travaux de Miles et Huberman (2003) et en tenant compte du journal de l'assistante de recherche. Mentionnons que l'analyse et la collecte des données se sont déroulées de façon concomitante, les données étant analysées au fur et à mesure. Cette façon de faire a nourri notre compréhension des besoins des femmes en cours de recherche et nous a permis de mieux les explorer lors des entrevues individuelles subséquentes et du *focus group*.

Le contenu des entrevues a fait l'objet d'une transcription. La première activité d'analyse a consisté à condenser les données. Pour ce faire, au moins deux membres de l'équipe de recherche ont lu chacune des entrevues dès leur transcription et ont procédé au codage des unités de sens. En réalisant l'analyse des premières entrevues, les membres de l'équipe de recherche ont élaboré une grille d'analyse qui incluait des catégories qui rendaient compte de la teneur des propos des participantes. Elles étaient regroupées, entre autres, selon qu'il était question du vécu émotif, des besoins de soutien psychologique, des besoins comblés et non comblés, et des suggestions de moyens pour répondre à leurs besoins. Cette grille a ensuite évolué au fil du processus d'analyse par l'ajout de catégories émergentes et par le regroupement des catégories de même nature en une structure unifiée et plus conceptuelle. Elle a conduit à la reconnaissance des composantes du vécu des participantes ainsi qu'à la définition des besoins et à une description initiale de ceux-ci.

L'équipe de recherche rédigeait des mémos relatant les questionnements de nature méthodologique ou analytique (p. ex., décision de regrouper tout ce qui concernait la relation au temps et la notion d'attente). La deuxième activité d'analyse a consisté à réviser les résultats de la condensation des données en explorant différentes façons de les présenter à l'aide de matrices conceptuelles et temporelles dans l'optique d'en dégager une compréhension plus globale et signifiante des besoins psychosociaux et informatifs des participantes. Ceci a permis l'émergence de catégories plus

conceptuelles comme la ligne de temps illustrant les moments d'attente pendant la période prédiagnostique et les représentations sociales du cancer du sein chez les participantes. La troisième activité d'analyse consistait à élaborer des conclusions et à les vérifier, ce qui a été fait tout au long de la recherche. Parmi les moyens utilisés à cette fin, mentionnons la coanalyse du contenu de 13 entrevues individuelles, la validation des résultats de l'analyse de l'ensemble des données au cours du *focus group* et les échanges réguliers au sein de l'équipe de recherche, qui nous ont permis d'améliorer notre compréhension des besoins des participantes.

2.7. Exemples de résultats

L'attente, l'anticipation et une gamme d'émotions fortes sont au cœur du vécu des femmes interviewées. Bien que l'expérience d'attente soit subjective et individuelle, elle est qualifiée de longue et souvent comparée à l'enfer. Pour les participantes, «*l'attente, c'est la pire chose*» (P-8, qui vient d'obtenir un rendez-vous pour une biopsie). Les propos suivants illustrent l'état émotionnel des participantes associé à l'attente: «*Cela me paraît une éternité… Je suis de plus en plus angoissée et j'ai très peur*» (P-12). Cela n'est pas étonnant considérant qu'elles voient le cancer du sein comme une menace sérieuse à la santé et à l'intégrité corporelle: «*Ce n'est pas la maladie cancer. C'est cette marque de quelque chose qui nous met en danger imminent*» (P-14). Le cancer du sein, «*c'est une tombe au bout…*» (P-18). En plus de ces représentations sombres du cancer du sein, les participantes sont habitées par d'autres représentations du cancer telles que la quête d'un sens à leur vie et à de multiples pertes.

C'est sur cette toile de fond qu'émergent les besoins des participantes en attente pour des examens complémentaires ou des résultats à la suite d'une mammographie révélant une anomalie. La connaissance de cette toile de fond sur laquelle émergent les besoins des participantes permet d'aller au-delà de la description d'une liste de besoins pour arriver à une compréhension plus holiste de leurs besoins sur les plans psychosocial et informatif. Quelques résultats (Doré, Gallagher, Saintonge, Grégoire et Hébert, 2013) sont présentés dans les paragraphes qui suivent afin d'illustrer à quel point les besoins sont teintés par l'expérience de l'attente et les émotions qui y sont associées.

La quasi-totalité des participantes expriment de nombreux besoins d'information non comblés. Elles ignorent tout du déroulement de la période prédiagnostique: les étapes du processus d'évaluation de leur situation, qui les renseignera sur les examens prévus, dans combien de temps les examens auront lieu, quand elles auront les résultats, etc. En outre, elles

déplorent le manque de clarté dans la transmission des résultats d'examens. À titre d'exemple, elles ne comprennent pas le sens des propos des professionnels de la santé lorsqu'ils se disent «*satisfaits*» (P-13) d'un résultat, qu'ils qualifient un résultat de «*nébuleux*» (P-28), ou qu'ils considèrent qu'elles sont «*à risque*» (P-26). Elles se questionnent aussi sur leur état de santé: les conséquences, les dangers, les risques associés à leur condition. Finalement, elles ne reçoivent pas d'information sur les ressources vers lesquelles se diriger «*pour poser des questions*» (P-12). À ce manque d'information, une participante réagit en mentionnant la perte de contrôle qu'il engendre: «*On ne peut pas avoir le contrôle à 100% parce qu'on ne sait pas, on est dans l'inconnu. Plus t'as de l'information, plus tu peux réagir. En tout cas, moi, j'aime mieux avoir la vérité… claire, nette et précise pour faire face aux choses*» (P-26).

La moitié des femmes interviewées affirme que leurs besoins de soutien psychologique ne sont pas comblés. Entre autres, les femmes déplorent le manque de soutien psychologique de la part du réseau de la santé, l'absence de personne-ressource vers qui se tourner, comme en font foi les propos d'une participante: «*Y a aucune ressource, y a personne qui te parle… bon j'ai eu des lettres. J'ai été informée par téléphone*» (P-12). Ces modes de communication ne répondent pas aux besoins ressentis par les participantes. On ne leur dit pas: «*Si vous avez besoin d'aide, appelez-nous. Quelqu'un est disponible pour vous parler*» (P-12). L'absence de services en ce sens font que les femmes ne se sentent pas soutenues pendant la phase prédiagnostique alors qu'une intense inquiétude les habite. Elles ont l'impression de manquer de soutien au moment où elles en ont le plus besoin, qu'elles sont en état de choc. Après un examen complémentaire (p. ex., échographie), elles sortent de la clinique «*le cœur gros*» (P-15) et apprécieraient avoir quelqu'un vers qui se tourner et une liste de ressources pouvant leur venir en aide.

Ces quelques exemples de résultats issus de la recherche descriptive interprétative renseignent les acteurs du réseau de la santé impliqués auprès des femmes ayant reçu un résultat anormal à leur mammographie de dépistage du cancer du sein sur les besoins de ces femmes. Forts de cette connaissance, ils peuvent réorienter leurs actions de façon à ce que l'expérience des femmes en pareille situation soit moins éprouvante et plus sereine en leur donnant accès au bon soutien, au bon moment, par une personne-ressource compétente.

CONCLUSION

En définitive, la recherche descriptive interprétative s'avère un dispositif de recherche utile à l'avancement des connaissances disciplinaires dans le domaine de la santé mentale. Elle permet la clarification et l'explication

de phénomènes rencontrés par les professionnels de la santé mentale. L'appropriation des connaissances qui émergent de ce type de recherche peut contribuer à une pratique professionnelle adaptée à la réalité des personnes visées par les interventions des professionnels. Aussi, une description interprétative peut ouvrir la voie à d'autres études, par exemple à une théorisation ancrée ou à une étude évaluative.

RÉFÉRENCES

CORBIN, J. et A. STRAUSS (2008). *Basics of Qualitative Research*, 3e éd., Thousand Oaks, Sage Publications.

CRABTREE, B.F. et W.L. MILLER (dir.) (1999). *Doing Qualitative Research*, 2e éd., Thousand Oaks, Sage Publications.

CRESWELL, J.W. (2013). *Qualitative Inquiry & Research Design. Choosing Among Five Approaches*, 3e éd., Thousand Oaks, Sage Publications.

DEMIR, F., Y.C. DONMEZ, E. OZSAKER et A. DIRAMALI (2008). « Patients' lived experiences of excisional breast biopsy : A phenomenological study », *Journal of Clinical Nursing*, vol. 17, no 6, p. 744-751.

DENZIN, N.K. et Y.S. LINCOLN (dir.) (2011). *The Sage Handbook of Qualitative Research*, Thousand Oaks, Sage Publications.

DORÉ, C., F. GALLAGHER, L. SAINTONGE, Y. GRÉGOIRE et M. HÉBERT (2013). « Breast cancer screening program : Experiences of Canadian women and their unmet needs », *Health Care for Women International*, vol. 34, no 1, p. 34-49.

DRAGESET, S. et T.C. LINDSTRØM (2003). « The mental health of women with suspected breast cancer : The relationship between social support, anxiety, coping and defence in maintaining mental health », *Journal of Psychiatric and Mental Health Nursing*, vol. 10, no 4, p. 401-409.

DRAGESET, S., T.C. LINDSTRØM et K. UNDERLID (2010). Coping with breast cancer : Between diagnosis and surgery. *Journal of advanced nursing,* vol. 66, no 1, p. 149-158. doi : 10.1111/j.1365-2648.2009.05210.x.

FORTIN, M.F. (2010). *Fondements et étapes du processus de recherche. Méthodes quantitatives et qualitatives*, 2e éd., Montréal, Chenelière Éducation.

GEOFFRION, P. (2009). « Le groupe de discussion », dans B. Gauthier (dir.), *Recherche sociale. De la problématique à la collecte des données*, 5e éd., Québec, Presses de l'Université du Québec, p. 391-414.

HECKMAN, B.D., E.B. FISHER, B. MONSEES, M. MERBAUM, S. RISTVEDT et C. BISHOP (2004). « Coping and anxiety in women recalled for additional diagnostic procedures following an abnormal screening mammogram », *Health Psychology*, vol. 23, no 1, p. 42-48.

JACCOUD, M. et R. MAYER (1997). « L'observation en situation et la recherche qualitative », dans J. Poupart, L.H. Groulx, J.P. Deslauriers, A. Laperrière, R. Mayer et A.P. Pires (dir.), *La recherche qualitative : enjeux épistémologiques et méthodologiques*, Boucherville, Gaëtan Morin, p. 211-249.

KITZINGER, J. (1994). « The methodology of focus groups : The importance of interactions between research participants », *Sociology of Health and Illness*, vol. 16, n° 1, p. 103-121.

KVALE, S. et S. BRINKMANN (2009). *InterViews. Learning the Craft of Qualitative Interviewing*, 2e éd., Thousand Oaks, Sage Publications.

LAPERRIÈRE, A. (1997). « Les critères de scientificité des méthodes qualitatives », dans J. Poupart, L.H. Groulx, J.P. Deslauriers, A. Laperrière, R. Mayer et A.P. Pires (dir.), *La recherche qualitative : enjeux épistémologiques et méthodologiques*, Boucherville, Gaëtan Morin, p. 365-389.

LASIUK, G.C., T. COMEAU et C. NEWBURN-COOK (2013). « Unexpected : An interpretative description of parental traumas' associated with preterm birth », *BMC Pregnancy and Childbirth*, vol. 13 (Suppl 1), <http://www.biomedcentral.com/1471-2393/13/S1/S13>.

LIAO, M.-N., M.-F. CHEN, S.-C. CHEN et P.-L. CHEN (2007). « Healthcare and support needs of women with suspected breast cancer », *Journal of Advanced Nursing*, vol. 60, n° 3, p. 289-298.

LIAO, M.-N., M.-F. CHEN, S.-C. CHEN et P.-L. CHEN (2008). « Uncertainty and anxiety during the diagnostic period for women with suspected breast cancer », *Cancer Nursing*, vol. 31, n° 4, p. 274-283.

LOISELLE, C.G., J. PROFETTO-MCGRATH, D.F. POLIT et C.T. BECK (2007). *Méthodes de recherche en sciences infirmières : approches quantitatives et qualitatives*, Montréal, ERPI.

MAHEU, C. et S. THORNE (2008). « Receiving inconclusive genetic test results : An interpretive description of BRCA1/2 experience », *Research in Nursing & Health*, vol. 31, p. 553-562.

MCCAUGHAN, E. et O. MCSORLEY (2007). « Consumers' and professionals' perceptions of a breast cancer review clinic », *Journal of Advanced Nursing*, vol. 60, n° 4, p. 419-426.

MILES, M.B. et M.A. HUBERMAN (2003). *Analyse des données qualitatives*, 2e éd., trad. de M.H. Rispal, Bruxelles, De Boeck.

MILES, M.B., M.A. HUBERMAN et J. SALDAÑA (2014). *Qualitative Data Analysis. A Methods Sourcebook*, 3e éd., Thousand Oaks, Sage Publications.

MILLER, W.L. et B.F. CRABTREE (2003). « Clinical research », dans N.K. Denzin et Y.S. Lincoln (dir.), *Strategies of Qualitative Inquiry*, Thousand Oaks, Sage Publications, p. 397-434.

O'REILLY, M. et N. PARKER (2013). « "Unsatisfactory saturation" : A critical exploration of the notion of saturated sample sizes in qualitative research », *Qualitative Research*, vol. 13, p. 190.

PAILLÉ, P. (1991). « Procédures systématiques pour l'élaboration d'un guide d'entrevue semi-directive : un modèle et une illustration », communication au 59e Congrès de l'Association canadienne-française pour l'avancement des sciences (21 au 24 mai), Sherbrooke, Québec.

PAILLÉ, P. et A. MUCCHIELLI (2010). *L'analyse qualitative en sciences humaines et sociales*, 2e éd., Paris, Armand Colin.

PELACCIA, T. et P. PAILLÉ (2010). « Les approches qualitatives : une invitation à l'innovation et à la découverte dans le champ de la recherche en pédagogie des sciences de la santé », *Pédagogie médicale*, vol. 10, n° 4, p. 293-304.

PIRES, A.P. (2007). « Échantillonnage et recherche qualitative : essai théorique et méthodologique », dans N.K. Denzin et Y.S. Lincoln (dir.), *Strategies of Qualitative Inquiry*, Thousand Oaks, Sage Publications, p. 113-169.

POUPART, J. (1997). « L'entretien de type qualitatif : considérations épistémologiques, théoriques et méthodologiques », dans J. Poupart, L.H. Groulx, J.P. Deslauriers, A. Laperrière, R. Mayer et A.P. Pires (dir.), *La recherche qualitative : enjeux épistémologiques et méthodologiques*, Boucherville, Gaëtan Morin, p. 173-210.

SANDELOWSKI, M. (1995). « Focus on qualitative methods. Sample size in qualitative research », *Research in Nursing and Health*, vol. 18, p. 178-183.

SANDELOWSKI, M. (2000). « Whatever happened to qualitative description ? », *Research in Nursing and Health*, vol. 23, p. 334-340.

SANDELOWSKI, M. (2008). « Foreword », dans S.E. Thorne (dir.), *Interpretive Description*, Walnut Creek, Left Coast Press, p. 11-14.

SANDELOWSKI, M. (2010). « What's in the name ? Qualitative description revisited », *Research in Nursing and Health*, vol. 33, p. 77-84.

SAVOIE-ZAJC, L. (2009). « L'entrevue semi-dirigée », dans B. Gauthier (dir.), *Recherche sociale. De la problématique à la collecte des données*, Québec, Presses de l'Université du Québec, p. 337-360.

STAKE, R.E. (1995). *The Art of Case Study Research*, Thousand Oaks, Sage Publications.

STATISTIQUE CANADA (2011). *Le cancer au Canada : cancers du poumon, du côlon et du rectum, du sein et de la prostate*, <http://www.statcan.gc.ca/pub/82-624-x/2011001/article/11596-fra.pdf>.

THORNE, S.E. (2008). *Interpretive Description*, Walnut Creek, Left Cost Press.

THORNE, S., A. CON, L. MCGINNESS, G. MCPHERSON et S.R. HARRIS (2004). « Health care communication issues in multiple sclerosis : An interpretive description », *Qualitative Health Research*, vol. 14, n° 1, p. 5-22.

THORNE, S., L. MCGINNESS, G. MCPHERSON, A. CON, M. CUNNINGHAM et S.R. HARRIS (2004). « Healthcare communication issues in fibromyalgia : An interpretive description », *Physiotherapy Canada*, vol. 56, n° 1, p. 31-38.

THORNE, S., J. OLIFFE, C. KIM-SING, T.G. HISLOP, K. STAJDUHAR, S.R. HARRIS, E.A. ARMSTRONG et V. OGLOV (2010). « Helpful communication during the diagnostic period : An interpretive description of patient preferences », *European Journal of Cancer Care*, vol. 19, p. 746-754.

THORNE, S., S. REIMER KIRKHAM et K. O'FLYNN-MAGEE (2004). « The analytic challenge in interpretive description », *International Journal of Qualitative Methods*, vol. 3, n° 1, article 1, <http://www.ualberta.ca/~iiqm/backissues/3_1/pdf/thorneetal.pdf>.

TOURÉ, E.H. (2010). « Réflexion épistémologique sur l'usage des *focus groups :* fondements scientifiques et problèmes de scientificité », *Recherches qualitatives*, vol. 29, n° 1, p. 5-27.

CHAPITRE 2

LA PHÉNOMÉNOLOGIE DE HUSSERL

Application de la méthode Investigation relationnelle *Caring* pour mieux comprendre l'expérience infirmière d'« être avec » la personne soignée en réadaptation

Louise O'Reilly
Chantal Cara

FORCES

- Elle s'attarde à recueillir la perspective des personnes vivant le phénomène à l'étude.
- Elle offre une perspective épistémologique.
- Elle permet l'obtention d'une description rigoureuse de la signification du phénomène étudié.

LIMITES

- Le nombre de participants est restreint.
- Les résultats reflètent uniquement la perspective des acteurs interrogés.
- La transférabilité des résultats de recherche est donc plus limitée.

En sciences humaines et de la santé, un nombre intéressant de méthodes phénoménologiques existe. À titre d'exemple, citons celles de Paterson et Zderad (1976), Colaizzi (1978), van Manen (1990), Ray (1991), Benner (1994), Giorgi, (1997) et Cara (1997, 1999), lesquelles méthodes découlent principalement de deux écoles de pensée, la phénoménologie transcendantale de Husserl et la phénoménologie interprétative et herméneutique de Heidegger. La phénoménologie transcendantale de Husserl constituera l'objet du présent chapitre.

La phénoménologie de Husserl (1970) offre une compréhension et une description en profondeur d'un phénomène d'intérêt, de même qu'elle rend possible l'émergence d'une signification (essence) pure et universelle de l'expérience étudiée. Afin de bien comprendre cette école de pensée, nous définirons la phénoménologie de Husserl et ces particularités épistémologiques. Puis suivra une présentation des méthodes phénoménologiques associées à l'école de pensée de Husserl, notamment celle de Cara (1997) – dont le *caring* et la dimension relationnelle correspondent à ses fondements philosophiques – ainsi que celle de Giorgi (1997) – méthode bien connue, très explicite et souvent utilisée. Pour clore cette première partie, nous discuterons des critères de scientificité en recherche qualitative. La deuxième partie du chapitre est une application concrète des étapes formant la méthode de Cara ; elle présente les dimensions méthodologiques de l'étude phénoménologique d'O'Reilly (2007), laquelle recherche visait à explorer la signification et la contribution de l'expérience d'« être avec » la personne soignée en contexte de réadaptation.

1. PHÉNOMÉNOLOGIE TRANSCENDANTALE DE HUSSERL

Edmund Husserl (1859-1938), philosophe allemand et d'abord mathématicien, est reconnu comme étant le père fondateur du mouvement phénoménologique (Giorgi, 1997 ; Spiegelberg, 1982). Pour ce mathématicien (Husserl, 1970), la science issue du paradigme positiviste, laquelle a pour but la vérité objective (mesurable), écarte les questions universelles telles que celles touchant la signification de l'existence humaine. Cela amène Husserl à réfléchir sur le point suivant : « Comment l'existence humaine peut-elle avoir une signification quelconque, si la science reconnaît comme seule vérité, ce qui est objectif ? » Ainsi, et tout au long de sa vie, Husserl se questionne et approfondit la phénoménologie, qu'il considère comme 1) une philosophie qui guide toute pensée dite scientifique puisqu'elle permet de prendre en compte la relation du chercheur avec son projet de recherche, rehaussant ainsi l'objectivité scientifique (*bracketing*) (Husserl, 1970), et 2) une psychologie qui procure un chemin à l'exploration de la

dimension spirituelle de l'existence humaine (Husserl, 1970, 1999). Pour Husserl (1970), la phénoménologie correspond à l'étude d'un phénomène d'intérêt dans le but de le décrire et de le comprendre. Cette phénoménologie se dit «épistémologique» puisqu'elle vise à décrire en détail et de façon rigoureuse la structure d'un phénomène. On dit alors que c'est une phénoménologie descriptive, laquelle contribue à une compréhension profonde de l'expérience étudiée. De même, la phénoménologie de Husserl est dite «transcendantale» puisque l'objectif final vise l'émergence d'une signification (essence) pure et universelle du phénomène, c'est-à-dire une essence qui expose le discours commun issu de l'ensemble des entrevues analysées.

La phénoménologie transcendantale de Husserl comprend quatre concepts épistémologiques majeurs. La conscience (*awareness*, *consciousness*), le premier d'entre eux, est essentielle dans la mesure où elle est «le moyen d'accès à tout ce qui se vit dans l'expérience, puisqu'il n'est rien qui puisse être dit ou à quoi on se rapporte qui n'inclut pas implicitement la conscience» (Giorgi, 1997, p. 343). Ainsi, la conscience, grâce à sa nature intrinsèque, contribue à donner une signification à l'expérience vécue. Le deuxième concept, l'intentionnalité (*intentionality*) ou la conscience intentionnelle (*intentional consciousness*), qui est un élément de la conscience, représente la relation d'intention indissociable entre le sujet (p. ex., participant de recherche), sa conscience (qui va donner un sens véritable au phénomène étudié) et l'objet de la phénoménologie (le phénomène à l'étude) (Giorgi, 1997; Husserl, 1999).

Le troisième concept épistémologique, la réduction phénoménologique, est constitué de deux attitudes, le *bracketing* et la réduction eidétique. Pour Husserl (1970, 1999), le *bracketing* (terme grec: *epochè*) désigne la reconnaissance et la mise entre parenthèses, par le chercheur, de ses valeurs, croyances, préjugés et connaissances théoriques relatives au phénomène à l'étude. Dans le même ordre d'idées, Giorgi (1997, p. 347) mentionne «qu'il faut mettre à l'écart ou rendre non influente toute connaissance passée en lien avec le phénomène sous investigation». Pour cet auteur, une attitude neuve est requise, c'est-à-dire une attitude de ne pas savoir (*unknowing*) afin d'accueillir et de ne pas juger les propos des participants. La deuxième attitude de la réduction phénoménologique se nomme la réduction eidétique (*épochè eidetic*). Elle correspond à un processus d'abstraction qui requiert de mettre de côté les faits particuliers ou individuels des témoignages pour tendre vers ce qui est universel (discours commun des entrevues). Grâce à ce processus d'abstraction, le chercheur parvient à faire émerger les structures essentielles (discours commun), lesquelles dévoilent la signification universelle du phénomène étudié. Cette signification est aussi qualifiée de pure puisqu'elle transcende («qui va au-delà») les discours, valeurs, croyances, préjugés et expériences personnels (Husserl, 1970, 1999).

Le dernier concept, l'intuition (*intuition*), s'intéresse au processus du chercheur pour décrire l'expérience quotidienne telle qu'elle est vécue par les participants, à partir d'une ouverture d'esprit et de l'utilisation de plusieurs modes de conscience (*multiple modes of awareness*) (Reeder, 1991). Aux yeux de Husserl (1970), c'est l'ouverture qui nous permet d'aller au-delà de l'expérience physique immédiate. Ainsi, parmi les divers modes de conscience, outre les cinq sens corporels, certains auteurs (Cara, 2002; Reeder, 1991) citent l'intuition intellectuelle, l'imagination, l'anticipation, la mémoire et les sentiments comme étant des exemples d'une conscience plus large pouvant être engagée pour décrire l'expérience.

1.1. But, question de recherche et échantillon

Comme mentionné préalablement, le but de la phénoménologie de Husserl consiste à décrire et à comprendre la signification de l'expérience pour la personne qui la vit quotidiennement (selon sa perception et sa perspective). De façon concrète, la question ou les questions de recherche découlent du but visé par la recherche. L'échantillon est de type raisonné et se compose de personnes ayant vécu l'expérience en lien avec le phénomène étudié et qui acceptent de partager leur expérience vécue (critères d'inclusion). Des critères d'exclusion peuvent être mentionnés et justifiés tels que ne pas comprendre une langue précise ni être capable de s'exprimer dans celle-ci. Selon le contexte et le but de l'étude, d'autres critères d'inclusion et d'exclusion peuvent s'ajouter. L'important est de viser un échantillon diversifié de participants (Lincoln et Guba, 1985). En effet, une plus grande diversité de participants permet de recueillir des propos divergents, ce qui contribue à enrichir la description du phénomène. Par conséquent, les résultats de recherche pourront mieux décrire le phénomène étudié, ce qui rehaussera la crédibilité des résultats de la recherche phénoménologique.

1.2. Collecte des données

Du côté de la collecte des données (guide d'entrevue et entrevues), celle-ci s'organise autour de la ou des questions de recherche. À partir de ces grandes questions de recherche, le chercheur aura la tâche de développer un guide d'entrevue dans lequel les sous-questions doivent être claires et leur ordre très bien établi. La clarté des sous-questions reste fondamentale afin d'obtenir des réponses pertinentes aux questions de recherche. L'ordre des questions demeure aussi important. Par exemple, il est suggéré de débuter par des questions plus simples, c'est-à-dire celles touchant l'expérience vécue de même que les perceptions et les sentiments en lien avec

le phénomène étudié, puis de terminer par des questions plus complexes, notamment celles en lien avec la signification du phénomène. Cette organisation logique de questions aide le participant à s'imprégner tranquillement du phénomène à l'étude. Finalement, pour tester la clarté et l'organisation des questions formant le guide d'entrevue, un moyen facile consiste à poser ces questions à des collègues ou à des experts en recherche qualitative.

Munhall (2012) recommande plus d'une entrevue, car, selon elle, la première entrevue sert à établir le lien de confiance entre le chercheur et le participant. Pour cette auteure, la réalisation d'entrevues phénoménologiques exhorte à se décentrer de soi pour être présent à l'autre. Sans cette condition, elle nous rappelle qu'il n'y a pas d'étude phénoménologique. Toujours selon Munhall, il est important, lors des entrevues, de préciser aux participants que c'est leur perception que nous recherchons, et qu'il n'y a donc pas de bonne ou de mauvaise réponse. Dans la même perspective que cette dernière, Paillé (2008) stipule que l'écoute et l'empathie, dans le cadre d'une collecte de données qualitatives, demeurent deux attitudes essentielles. Paillé définit la première comme étant « l'écoute initiale complète et totale des témoignages pour ce qu'ils ont à nous apprendre, avant que nous soyons tentés de les faire parler. Cela consiste à donner la parole, d'accorder de la valeur à l'expérience de l'autre. Les propos des personnes ne sont pas la réalité du chercheur. C'est pour cette raison qu'il doit donner la parole aux autres puisque c'est eux qui connaissent leur réalité » (p. 86). Il décrit ensuite l'empathie de cette manière :

> Éviter un jugement hâtif ou interprétation puisqu'on ne veut pas parler de soi, mais bien des personnes que l'on a questionnées. Toute notre attention est portée vers les *verbatim* d'autrui. Une valeur de vérité leur est momentanément accordée. Il s'agit d'honorer la situation observée ou le témoignage rendu, d'accorder du crédit à ce qui a été exprimé, se laisser toucher, lâcher prise, par rapport à nos catégories interprétatives impératives. C'est de voir, penser et comprendre autrement. Se laisser transformer (p. 88).

2. MÉTHODE PHÉNOMÉNOLOGIQUE SCIENTIFIQUE DE GIORGI

La méthode phénoménologique applicable aux sciences humaines et de la santé d'Amedeo Giorgi (1997) découle de l'école de pensée de Husserl. De ce fait, la réduction phénoménologique de Husserl, définit précédemment, demeure centrale à sa méthode phénoménologique. La méthode phénoménologique de Giorgi se compose de cinq étapes, que nous détaillons dans les lignes qui suivent.

2.1. Collecte des données verbales

Cette étape concerne la collecte des témoignages grâce à des entrevues organisées autour de questions larges et ouvertes, lesquelles entrevues seront enregistrées et transcrites ultérieurement en comptes rendus intégraux (*verbatim*). Comme indiqué par Giorgi (1997), le but visé est « une description concrète et détaillée de l'expérience et des actes du sujet, qui soit aussi fidèle que possible à ce qui est arrivé tel qu'il l'a vécu » (p. 353).

2.2. Lecture des données

Elle rappelle l'importance de lire attentivement et au complet l'ensemble des données recueillies avant de commencer l'étape de l'analyse.

2.3. Division des données en unités de signification

Elle vise la constitution d'unités de signification appropriées. Ainsi et au fur et à mesure d'une relecture attentive des témoignages, il se peut que le chercheur observe un changement de sens dans le texte et fasse ainsi émerger de nouvelles unités de signification. Comme le mentionne Giorgi (1997), « cette étape permet au chercheur de rester plus proche des données que s'il tentait de les appréhender dans leur totalité » (p. 354).

2.4. Organisation et énonciation des données brutes dans le langage de la discipline

Chaque unité de signification générée à l'étape précédente sera à nouveau examinée et décrite de façon à rendre plus explicite sa valeur respective au regard de la discipline. Le rôle de la variation libre et imaginaire est majeur à cette étape-ci, puisqu'elle permet d'appréhender les liens entre les unités de signification relevées et les concepts disciplinaires pertinents.

2.5. Synthèse des résultats

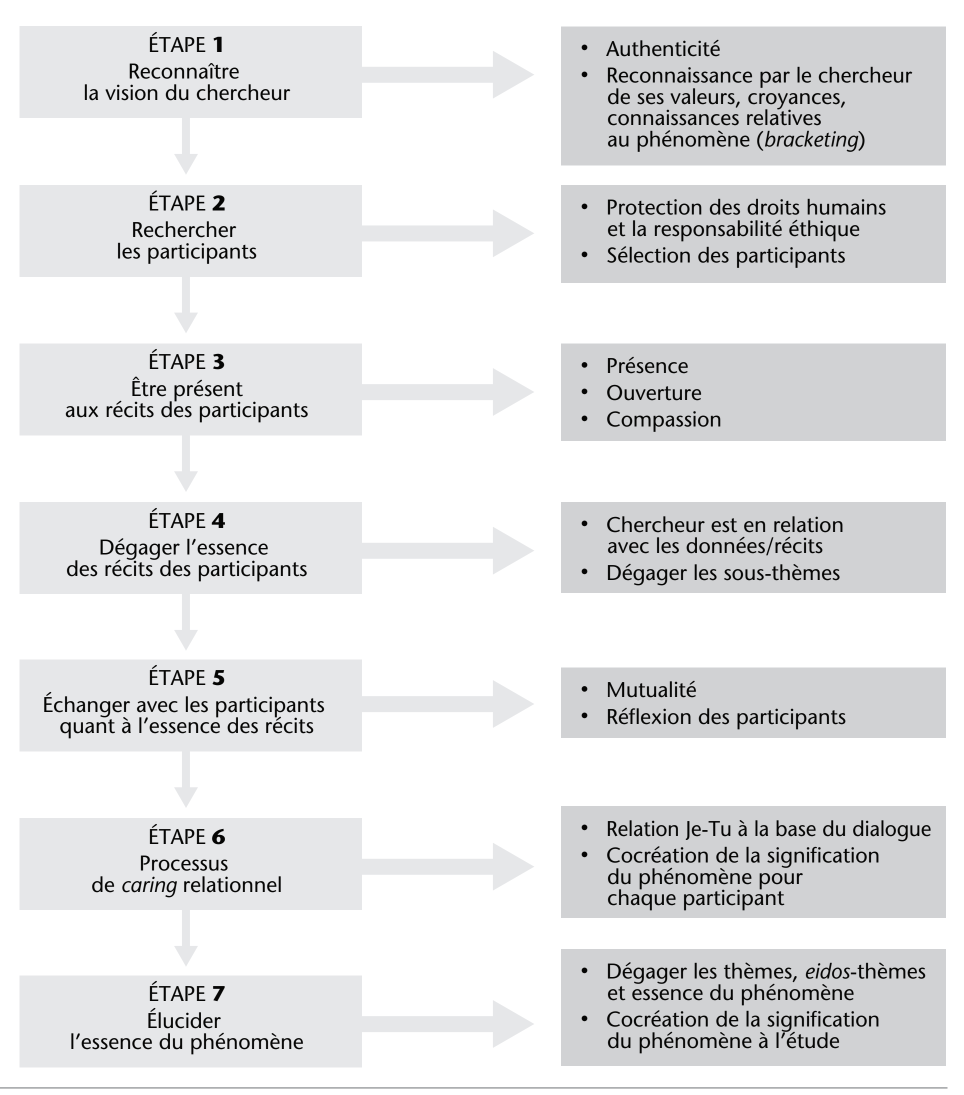
Cette dernière étape vise à examiner, à nouveau, les unités de signification transformées en langage de la discipline afin de distinguer lesquelles d'entre elles sont essentielles dans le cadre du phénomène à l'étude et lesquelles ne le sont pas. Le but est de décrire la structure essentielle du phénomène étudié. Ce travail sera facilité par l'utilisation de la méthode de la variation libre et imaginaire.

3. INVESTIGATION RELATIONNELLE *CARING* DE CARA

L'Investigation relationnelle *Caring* (Relational Caring Inquiry) (Cara, 1997, 1999, 2002) (figure 2.1) est une méthode phénoménologique qualifiée de relationnelle, dialogique, transformative et *caring*, qui vise la description et la compréhension des phénomènes d'intérêt issus de la discipline infirmière et de toutes les autres disciplines de la santé. Cara (1997) considère que le *caring* constitue l'essence de la discipline infirmière et qu'à ce titre, il se doit de pénétrer tous les champs d'activités qui y sont liés, y compris celui de la recherche. Ainsi, et jugeant que la dimension du *caring* se retrouve peu investie dans l'école de pensée de Husserl, Cara (1997) développe pendant ses études doctorales sa méthode, qu'elle nomme « Relational Caring Inquiry ». Les assises de cette méthode découlent d'écrits liés à la philosophie du *caring* (Buber, 1970 ; Gadow, 1994 ; Watson, 1985, 1988) et à l'école de pensée de Husserl (Husserl, 1970 ; Ray, 1991 ; Reeder, 1991). Ainsi, les fondements philosophiques de l'Investigation relationnelle *Caring* (Cara, 1997, 1999) correspondent aux caractéristiques associées au *caring*, telles que la présence authentique à chaque participant qui partage sa signification associée à son expérience du phénomène à l'étude, de même que, lors de l'analyse et de l'interprétation des *verbatim*, l'ouverture d'esprit à ce qui est exprimé, l'authenticité de soi et l'engagement. Le développement de la perspective relationnelle de cette méthode fut influencé par les réflexions de Buber (1970) sur les relations « Je-Tu » et « Je-Il ». Selon les écrits de ce philosophe, la relation « Je-Tu » se distingue de celle nommée « Je-Il » par l'implication de la globalité de la personne, de même que par la qualité de sa présence. Par ses assises issues du *caring*, cette méthode phénoménologique de Cara (1997) contribue à considérer la recherche comme un processus relationnel humain.

À ce jour, cette méthode fut retenue dans le cadre d'études (devis qualitatif et mixte) de la discipline infirmière qui s'intéressaient à la signification des phénomènes suivants : 1) l'influence de l'infirmière gestionnaire sur les habiletés des infirmières à s'engager dans leur pratique de *caring* (Cara, 1997) ; 2) l'expérience d'« être avec » la personne soignée selon la perspective des infirmières (O'Reilly, 2007) et celle des patients (O'Reilly *et al.*, 2010) ; 3) les changements perçus dans la pratique de soin humaniste à la suite de l'intervention éducative (O'Reilly et Cara, 2011) ; et 4) la qualité de vie au travail des cadres gestionnaires infirmiers de premier niveau œuvrant en établissement de santé (Brousseau, Cara et Blais, 2013). L'Investigation relationnelle *Caring* se compose de sept étapes interreliées et interdépendantes, lesquelles sont décrites dans les prochaines lignes.

Figure 2.1.
Investigation relationnelle *Caring*



Source : Cara, 2002 ; traduction libre.

3.1. Reconnaître la vision du chercheur

Il est question ici du *bracketing* que le chercheur doit réaliser afin de reconnaître ses valeurs, croyances, préjugés et connaissances reliés au phénomène à l'étude.

3.2. Rechercher les participants

Cette étape concerne les considérations éthiques de l'étude, le milieu de la recherche et la sélection des participants. La dimension éthique comprend l'obtention d'un certificat éthique, la signature du consentement écrit à la première entrevue, la protection de l'anonymat des participants, de même que la planification d'une ressource d'aide dans le cas où un participant ressentirait le besoin de parler à une personne qualifiée. Dans le cadre de l'autre volet, qui correspond au milieu de la recherche et aux participants, le chercheur doit offrir une description du milieu où se concrétisera la recherche, de même qu'énoncer les critères d'inclusion et d'exclusion pour le recrutement des participants.

3.3. Être présent aux récits des participants

Cette étape convie le chercheur à démontrer une présence authentique, de l'ouverture et de la compassion dans le but d'être réceptif et de mieux comprendre les expériences partagées de chaque participant à la recherche.

3.4. Dégager l'essence des récits des participants

Cela correspond à l'élucidation de l'essence des récits des participants (figure 2.2). Plus concrètement, il est question de la transcription de chaque entrevue, de la transformation de ces derniers en récit synthétique et de l'analyse. Pour Cara (1997), l'ensemble du processus d'analyse (verbatim et journal de bord) et d'interprétation des données demeure encadré par la réduction phénoménologique de Husserl (1970). Le processus d'analyse est avant tout circulaire, puisqu'il invite le chercheur à de multiples lectures des *verbatim*, à une ouverture d'esprit à ce qui est exprimé par chaque participant et à une réflexion pour parvenir à une compréhension profonde des opinions révélées.

3.5. Échanger avec les participants quant à l'essence des récits

Cette étape vise la mutualité entre le chercheur et chacun des participants à la recherche. Le but visé est d'aider chaque participant à clarifier le phénomène à l'étude. Pour y parvenir, chaque participant reçoit une copie du récit synthétique de la première entrevue, de même qu'une copie de l'analyse et de l'interprétation du récit. Il leur est demandé de lire les documents et d'apporter les corrections appropriées, lesquelles seront partagées avec le chercheur lors de la deuxième entrevue.

3.6. Processus de *caring* relationnel

Il cible les deux dernières entrevues (deuxième et troisième) réalisées avec le participant et se distingue par le dialogue entre le participant et le chercheur. Par ce dialogue, le chercheur souhaite obtenir le point de vue respectif des participants quant à son travail d'interprétation des données (récit synthétique et analyse des verbatim), c'est-à-dire une cocréation de la signification du phénomène pour chaque participant.

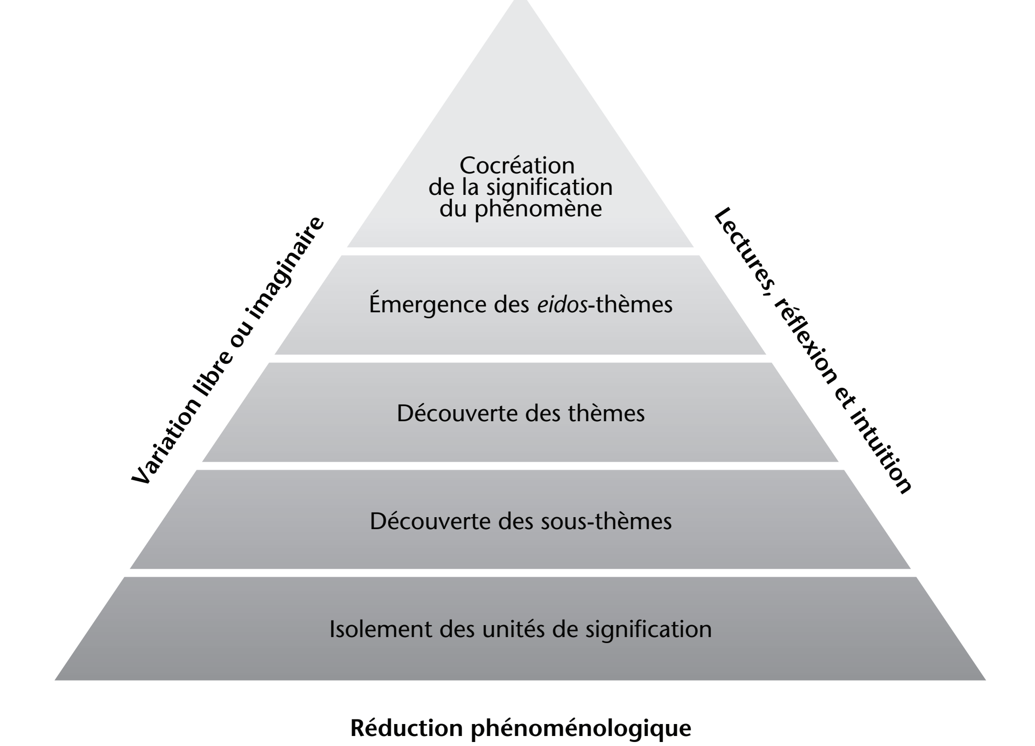
3.7. Élucider l'essence du phénomène

Pour accéder à l'essence, Cara (1997) recommande d'entreprendre une réduction eidétique, qui consiste à aller au-delà des faits particuliers ou individuels (de chaque récit) pour favoriser l'émergence de la signification universelle du phénomène à l'étude. Pour accomplir ce processus d'abstraction (réduction eidétique), certains auteurs (Cara, 2002; Reeder, 1991) recommandent l'utilisation de divers modes de conscience, notamment l'intuition intellectuelle, l'imagination, l'anticipation, la mémoire, les sentiments, l'expérience, la variation libre ou imaginaire, et certains sens corporels tels que la vue, le toucher et l'ouïe. Selon Cara (1997), la variation libre ou imaginaire facilite le regroupement des données, puisqu'elle consiste à remettre en question la place de chaque élément formant un regroupement de données. Par exemple, si je fais l'exercice de retirer un élément d'un regroupement, je dois me demander si cela change la nature du regroupement. Si la réponse est oui, cela signifie que l'élément est bel et bien dans le bon regroupement. Si, au contraire, le retrait de cet élément ne change en rien la nature du regroupement, cela indique que l'élément ciblé n'est pas dans le bon regroupement.

4. CRITÈRES DE SCIENTIFICITÉ EN RECHERCHE QUALITATIVE

Certains auteurs (Guba et Lincoln, 1981; Lincoln et Guba, 1985; Sandelowski, 1986; Whittemore, Chase et Mandle, 2001) réclament que l'authenticité, la crédibilité, la critique, l'intégrité et la transférabilité demeurent les principaux critères de scientificité en recherche qualitative. L'*authenticité*, aussi nommée «*confirmability*» (Sandelowski, 1986), permet de préciser si les résultats correspondent bien à la signification ou à l'expérience telle qu'elle a été décrite par chacun des participants (Cara, 2002; Sandelowski, 1986; Whittemore, Chase et Mandle, 2001). Des stratégies sont proposées telles que le respect de la réduction phénoménologique (*bracketing* ou *epochè*), les multiples lectures des entrevues ou récits, l'ouverture d'esprit et l'utilisation de divers modes de conscience par le chercheur, l'enregistrement

Figure 2.2.
Élucidation de l'essence du phénomène



Source : Cara, 2002 ; traduction libre.

des entrevues, la réalisation de la deuxième et de la troisième entrevues, ainsi que la tenue d'un journal de bord (Cara, 2002; Lincoln et Guba, 1985). Le critère de *crédibilité* s'observe lorsque les résultats de recherche décrivent vraiment le phénomène à l'étude (Cara, 2002; Lincoln et Guba, 1985; Whittemore, Chase et Mandle, 2001) et que les lecteurs et experts reconnaissent l'expérience décrite à la lecture de l'étude et de ses résultats (Sandelowski, 1986). Le choix diversifié des participants à la recherche, les entrevues réalisées jusqu'à la redondance des données, la variation libre et imaginative et la réduction eidétique, de même que la reconnaissance du phénomène par les lecteurs et experts (Cara, 2002; Sandelowski, 1986) permettent de rehausser la crédibilité des résultats. La *critique*, aussi appelée «*auditability*» (Sandelowski, 1986) ou «*dependability*» (Lincoln et Guba, 1985), désigne l'attitude critique constante du chercheur afin d'éviter d'être biaisé (Cara, 2002; Whittemore, Chase et Mandle, 2001). Une étude et ses

résultats seront jugés consistants lorsqu'un autre chercheur pourra arriver à des conclusions pareilles ou semblables, mais non contradictoires (Guba et Lincoln, 1981; Sandelowski, 1986). Certains moyens sont suggérés (Cara, 2002) tels que le respect de la réduction phénoménologique tout au long du processus de la recherche, l'utilisation de la variation libre et imaginative, la tenue d'un journal de bord, de même que les discussions avec des personnes considérées comme expertes dans le domaine à l'étude. L'*intégrité* illustre la préoccupation du chercheur de valider ses interprétations en s'appuyant sur les données (Cara, 2002; Whittemore, Chase et Mandle, 2001), par exemple en évitant de formuler l'essence du phénomène de façon prématurée, en atteignant la redondance des résultats avant de clôturer les entrevues, en utilisant une deuxième et troisième entrevue qui permettent au chercheur de valider son interprétation des *verbatim* (Cara, 2002; Whittemore, Chase et Mandle, 2001). Le dernier critère de scientificité, la *transférabilité*, aussi nommée «*fittingness*» (Sandelowski, 1986), qui correspond à des résultats de recherche qui peuvent être transférés à des situations similaires (Cara, 2002; Sandelowski, 1986), s'obtient en offrant une description détaillée des résultats de recherche (Lincoln et Guba, 1985).

5. APPLICATION DE LA MÉTHODE PHÉNOMÉNOLOGIQUE INVESTIGATION RELATIONNELLE *CARING*

Dans son étude doctorale, O'Reilly (2007) a cherché à mieux comprendre la signification du concept «être avec», lequel concept définit la nature du soin infirmier issue du paradigme de la transformation. Plusieurs auteurs de la discipline infirmière réclament la valeur inestimable de la relation de *caring* pour la clientèle de réadaptation, dont l'expérience d'«être avec» la personne soignante s'avère centrale. En dépit de cette importance, la recherche concernant les bienfaits thérapeutiques de la relation de *caring*, notamment de l'expérience d'«être avec» la personne soignée pour la clientèle nécessitant des soins et services de réadaptation, demeure un domaine peu exploité. Dans ce contexte, le but de la recherche phénoménologique d'O'Reilly visait essentiellement à explorer et à comprendre, auprès d'infirmières œuvrant dans un milieu clinique de réadaptation, la signification de l'expérience d'«être avec» la personne soignée, de même que leur perception de la contribution de cette expérience à la réadaptation de la personne soignée. Découlant de ce but, les deux grandes questions de recherche étaient: 1) Quelle est la signification de l'expérience d'«être avec» la personne soignée, selon la perception d'infirmières œuvrant en

réadaptation ? 2) Comment l'expérience d'« être avec » la personne soignée contribue-t-elle à la réadaptation, selon la perception d'infirmières œuvrant en réadaptation ?

L'Investigation relationnelle *Caring* (Cara, 1997), méthode phénoménologique retenue dans le cadre de cette étude, est en cohérence avec la toile de fond disciplinaire choisie, soit la théorie du « *human caring* » de Watson (2008). En effet, cette théoricienne recommande une approche phénoménologique-existentielle afin de comprendre la signification et l'expérience vécue par la personne, telle que celle-ci les décrit. À cet élément de justification ajoutons que les assises philosophiques de la méthode phénoménologique de Cara, celles en lien avec l'ontologie du *caring* et de la recherche, viennent contribuer positivement à la cohérence de ce choix méthodologique. Regardons maintenant chacune des sept étapes formant l'Investigation relationnelle *Caring*.

5.1. Reconnaître la vision du chercheur

L'ensemble des présuppositions (*bracketing*) de la chercheuse (p. ex., soin infirmier, phénomène d'« être avec ») a été reconnu et mis sur papier. Par exemple, la reconnaissance que le soin infirmier est de nature relationnelle, de même que le fait que la présence émotive permet d'« être avec » soi ou l'autre illustrent le point de vue de la chercheuse envers la nature respective du soin infirmier et du phénomène d'« être avec ». De plus, cette étape a aussi permis à la chercheuse de dessiner une conceptualisation du phénomène d'« être avec » la personne soignée à l'intérieur de la relation de *caring*. Ainsi et tout au long du processus de l'étude, la chercheuse a veillé à mettre entre parenthèses ses présuppositions.

5.2. Rechercher les participants

La chercheuse a soumis le projet de recherche aux deux comités éthiques de la recherche de chacun des deux centres de réadaptation visés par l'étude. Du côté de la protection des droits humains, les infirmières participant à la recherche avaient la possibilité de se retirer à n'importe quel moment du processus de la recherche. Aussi, la direction des soins infirmiers avait fait parvenir aux infirmières répondant aux critères d'inclusion les documents explicatifs de la recherche (résumé de la recherche, consentement écrit, feuillet explicatif). Dans le cadre d'une recherche qualitative, il est difficile de prédire avec exactitude le nombre de participants qui feront partie de l'étude. À ce sujet, Benner (1994) mentionne que de 5 à 25 participants

peuvent être nécessaires, tout en précisant la nécessité de poursuivre les entrevues jusqu'à ce qu'il y ait redondance des données. Dans ce cas-ci, la redondance a été atteinte avec 17 participants. D'autres stratégies ont été mises de l'avant pour protéger les droits humains, notamment celui d'utiliser un numéro et un pseudonyme pour chaque participant, de même que la planification d'une ressource (programme d'aide aux employés) au cas où un participant aurait eu besoin de parler à une personne qualifiée. Rappelons qu'en phénoménologie (Munhall, 2012), les études peuvent porter sur des sujets délicats, de même qu'exiger plus d'une entrevue. Dans ce contexte, on peut comprendre que le participant puisse ressentir le besoin de parler à une personne qualifiée.

5.3. Être présent aux récits des participants

La première rencontre (60 minutes) comprenait le questionnaire sociodémographique et la première question de recherche, à savoir « Quelle est la signification de l'expérience d'"être avec" la personne soignée, selon la perception d'infirmières œuvrant en réadaptation ? ». Chaque participant a été invité à partager ses réflexions personnelles, de même que certaines expériences, situations, actions, perceptions et émotions se rapportant à cette première question. Les sous-questions posées en lien avec la première grande question de recherche étaient : 1) Racontez-moi une expérience de travail dans laquelle vous avez ressenti « être avec » votre patient ? 2) Qu'est-ce que vous vous dites à vous-même au sujet de votre expérience d'« être avec » le patient ? 3) Quels sont vos sentiments en lien avec votre expérience d'« être avec » votre patient ? 4) Que signifie pour vous « être avec » votre patient ? À cette même étape, la chercheuse a apporté une attention spéciale à la qualité de sa présence auprès de chacun des participants. Pour y accéder, la pratique d'exercices de centration quelques minutes avant chaque entrevue a été intégrée au processus de la recherche. Les notes personnelles de la chercheuse ont été consignées consciencieusement dans un journal de bord tout de suite après chaque entrevue. À la fin de la première entrevue, la chercheuse a remercié chaque participant pour le partage et a mentionné qu'une deuxième entrevue devait être planifiée afin de discuter de l'analyse et de l'interprétation des *verbatim* associées à la première question de recherche.

5.4. Dégager l'essence des récits des participants

Une attention au *bracketing* de Husserl (1970, 1999) a été portée tout au long de la transcription des entrevues sous forme de textes (*verbatim*), qui a été réalisée en gardant le plus possible la perspective des participants (en utilisant leurs propres mots). Par la suite et après plusieurs lectures, la chercheuse a transformé le texte en récit synthétique (quelques pages). À ce moment, chaque récit synthétique a été analysé afin d'isoler les unités de signification (premier niveau de réflexion), c'est-à-dire les idées spécifiques au phénomène à l'étude. Puis, les unités de significations ont été regroupées en sous-thèmes (deuxième niveau de réflexion) qui correspondaient à des éléments descriptifs reflétant le plus possible le contenu des entrevues des participants.

5.5. Échanger avec les participants quant à l'essence des récits

Avant de procéder à la deuxième entrevue, chaque participant a reçu une copie de son récit synthétique (première question de recherche) analysé. Il leur a été demandé de lire les documents et d'apporter les corrections appropriées avant la deuxième entrevue, qui se déroulait environ un mois après la première.

5.6. Processus de *caring* relationnel

Cette étape correspond à la deuxième (45 minutes) et à la troisième entrevue (10 minutes).

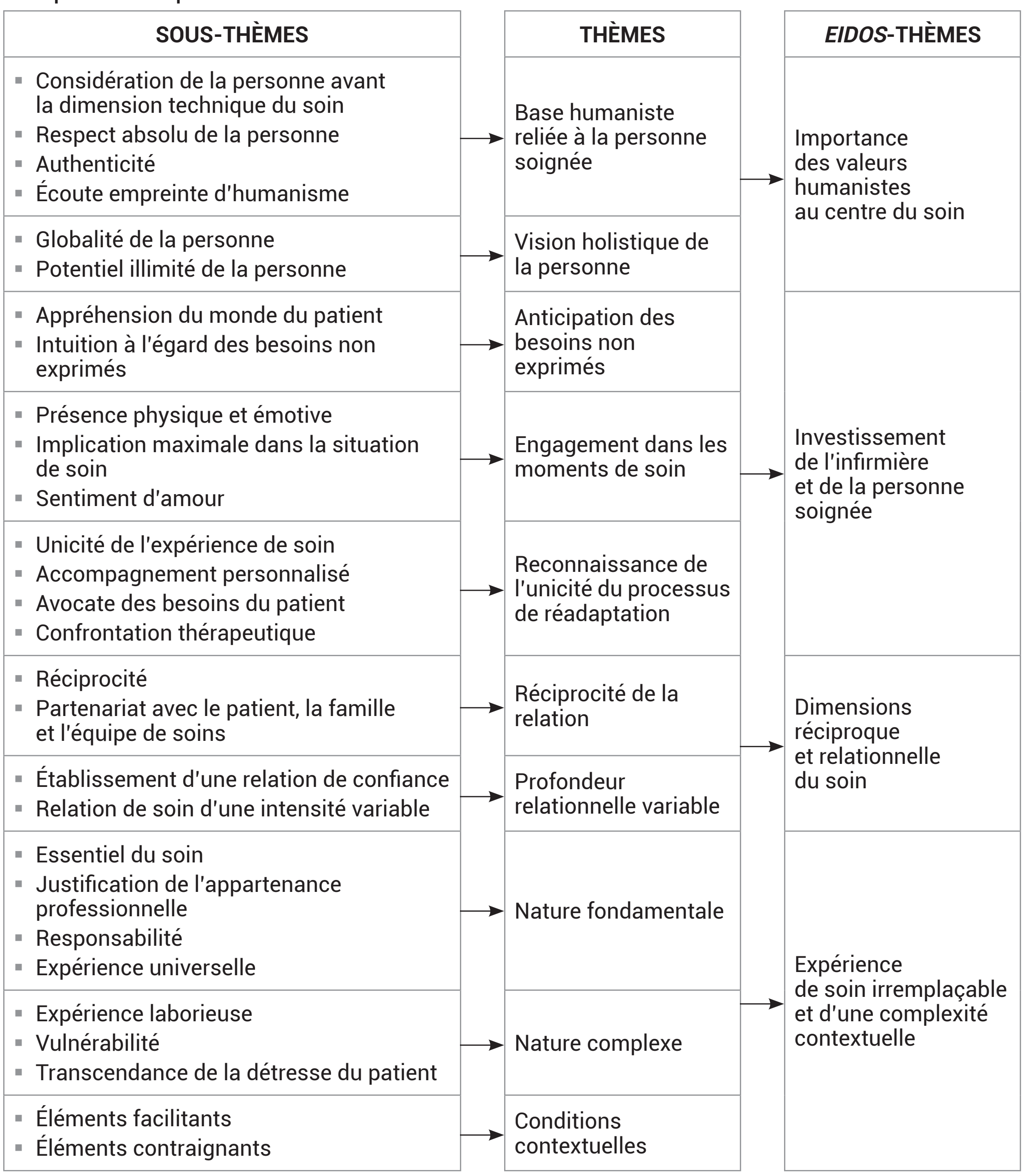
Au début de la deuxième entrevue, la chercheuse a demandé à chaque participant de partager son accord (ou son désaccord) en lien avec les unités de signification et les sous-thèmes retenus. Dans un deuxième temps, chaque participant a été invité à répondre à la seconde question de recherche, à savoir « Comment l'expérience d'être avec la personne soignée contribue-t-elle à la réadaptation ? ». Puis, l'analyse et l'interprétation de ces nouveaux *verbatim* ont été réalisées par un retour aux étapes antérieures, soit les étapes 3, 4, 5, et 6 (voir la figure 2.1). Avant la troisième entrevue (relance téléphonique), une copie du récit analysé (deuxième question de recherche) a été envoyée à chaque participant. Tout comme pour la première question de recherche, il leur était demandé de lire les documents et d'apporter les corrections appropriées avant cette dernière rencontre, soit un mois après la deuxième entrevue. Finalement, la relance téléphonique, qui a été effectuée auprès de chaque participant à leur convenance, termine cette sixième étape. Tout comme à la deuxième entrevue, la chercheuse a demandé à

chaque participant de partager son accord (ou son désaccord) en lien avec les unités de signification et les sous-thèmes retenus. À la fin, la chercheuse a exprimé sa gratitude pour la collaboration et la contribution de chaque participant au projet de recherche. Ils ont été informés qu'une copie du résumé du projet de recherche leur sera acheminée, lorsque la recherche sera terminée.

5.7. Élucider l'essence du phénomène

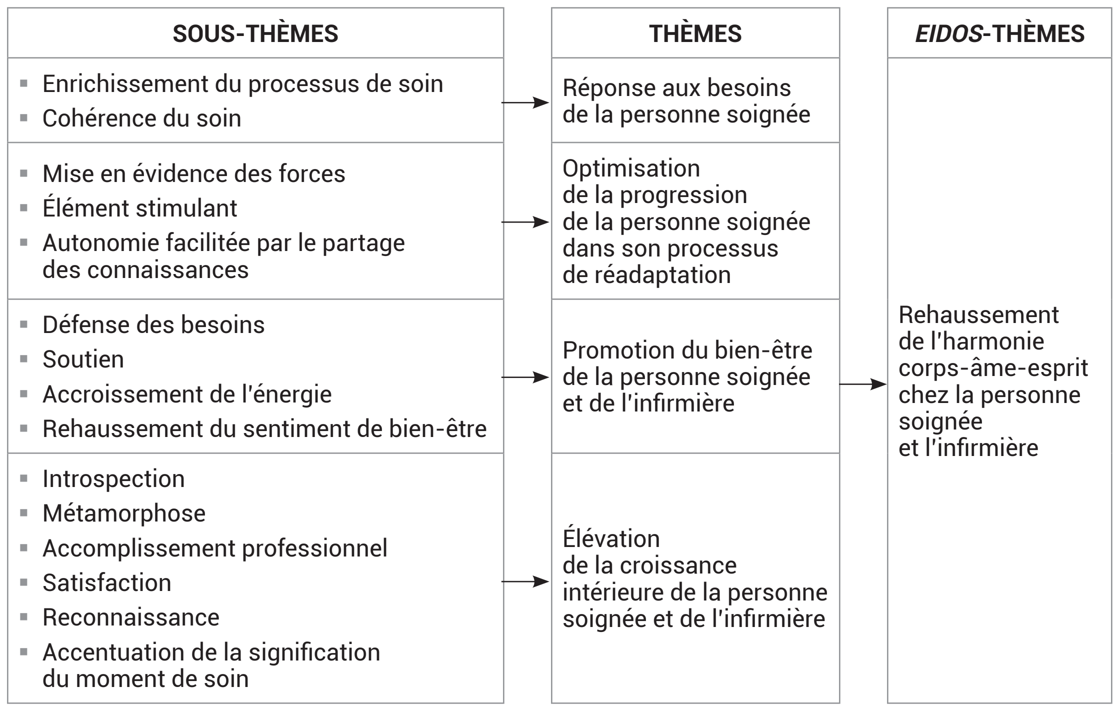
Pour parvenir à l'essence, les sous-thèmes retenus préalablement ont été rassemblés en thèmes (troisième niveau de réflexion) selon la convergence des idées qu'ils dégageaient. Puis, l'analyse approfondie des thèmes jumelée à l'utilisation de la variation libre et imaginaire ont favorisé l'émergence des *eidos*-thèmes (quatrième niveau de réflexion). Les *eidos*-thèmes retenus ont permis l'émergence de l'essence universelle (structure essentielle) du phénomène à l'étude (cinquième niveau de réflexion). Le processus d'abstraction, associé aux *verbatim* liés à la première et à la deuxième question de recherche, se retrouve illustré dans les tableaux 2.1 et 2.2. Cette forme demeure une façon pertinente de présenter les résultats puisqu'elle offre clarté et concision. Finalement, le processus d'analyse et d'interprétation des données a permis la découverte de cinq *eidos*-thèmes rattachés aux deux questions de recherche (quatre pour la première question de recherche, un pour la deuxième), ce qui a mené à l'émergence de la signification universelle de l'expérience d'« être avec » la personne soignée, laquelle est symbolisée par cette phrase : « Rencontre humaine profonde, thérapeutique et transformatrice ». Les prochaines lignes dévoilent les liens entre chaque terme formant l'essence et les *eidos*-thèmes, thèmes et sous-thèmes ayant émergé.

Tableau 2.1.
Émergence des sous-thèmes, thèmes et *eidos*-thèmes reliés à la première question de recherche



SOUS-THÈMES	THÈMES	*EIDOS*-THÈMES
▪ Considération de la personne avant la dimension technique du soin ▪ Respect absolu de la personne ▪ Authenticité ▪ Écoute empreinte d'humanisme	Base humaniste reliée à la personne soignée	Importance des valeurs humanistes au centre du soin
▪ Globalité de la personne ▪ Potentiel illimité de la personne	Vision holistique de la personne	
▪ Appréhension du monde du patient ▪ Intuition à l'égard des besoins non exprimés	Anticipation des besoins non exprimés	Investissement de l'infirmière et de la personne soignée
▪ Présence physique et émotive ▪ Implication maximale dans la situation de soin ▪ Sentiment d'amour	Engagement dans les moments de soin	
▪ Unicité de l'expérience de soin ▪ Accompagnement personnalisé ▪ Avocate des besoins du patient ▪ Confrontation thérapeutique	Reconnaissance de l'unicité du processus de réadaptation	
▪ Réciprocité ▪ Partenariat avec le patient, la famille et l'équipe de soins	Réciprocité de la relation	Dimensions réciproque et relationnelle du soin
▪ Établissement d'une relation de confiance ▪ Relation de soin d'une intensité variable	Profondeur relationnelle variable	
▪ Essentiel du soin ▪ Justification de l'appartenance professionnelle ▪ Responsabilité ▪ Expérience universelle	Nature fondamentale	Expérience de soin irremplaçable et d'une complexité contextuelle
▪ Expérience laborieuse ▪ Vulnérabilité ▪ Transcendance de la détresse du patient	Nature complexe	
▪ Éléments facilitants ▪ Éléments contraignants	Conditions contextuelles	

Tableau 2.2.
Émergence des sous-thèmes, thèmes et *eidos*-thèmes reliés à la deuxième question de recherche



SOUS-THÈMES		THÈMES		*EIDOS*-THÈMES
▪ Enrichissement du processus de soin ▪ Cohérence du soin	→	Réponse aux besoins de la personne soignée	→	Rehaussement de l'harmonie corps-âme-esprit chez la personne soignée et l'infirmière
▪ Mise en évidence des forces ▪ Élément stimulant ▪ Autonomie facilitée par le partage des connaissances	→	Optimisation de la progression de la personne soignée dans son processus de réadaptation		
▪ Défense des besoins ▪ Soutien ▪ Accroissement de l'énergie ▪ Rehaussement du sentiment de bien-être	→	Promotion du bien-être de la personne soignée et de l'infirmière		
▪ Introspection ▪ Métamorphose ▪ Accomplissement professionnel ▪ Satisfaction ▪ Reconnaissance ▪ Accentuation de la signification du moment de soin	→	Élévation de la croissance intérieure de la personne soignée et de l'infirmière		

Dans un premier temps, pour que se réalise cette expérience, il faut une *rencontre* entre deux personnes concernées, l'infirmière et la personne soignée. Voici un témoignage de Grâce, qui parle de la relation de réciprocité : « *Je pense qu'"être avec" mon patient représente une relation tissée entre la personne que je soigne et moi. C'est un échange qui va dans les deux sens. Cette relation commence dès les premiers échanges surtout en réadaptation, c'est un travail d'équipe. Cet échange permet à chacun de s'exprimer, par la voix ou autre chose.* » Or, pour que se concrétise cette rencontre, chacune de ces personnes doit, préalablement, accepter de prendre part à cette expérience de soin. Ainsi, cet engagement rendra possible le développement d'une relation de confiance dont la profondeur variera d'une personne soignée à l'autre sous l'effet de certains éléments de l'environnement des personnes convoitées par l'expérience d'être avec la personne soignée ou bien, impliquée dans l'expérience d'« être avec » la personne soignée. Deuxièmement, cette rencontre se qualifie d'*humaine* parce qu'elle accorde une valeur inestimable à la dimension humaine de chaque personne impliquée dans cette expérience de soin. Un témoignage de Dolorès illustre une des valeurs

humanistes, celle du respect : « […] *je savais qu'il y avait quelque chose, mais, avec lui, il ne fallait pas que j'aille questionner, il fallait que je l'attende ; si ce n'était pas ce soir-là, c'était un autre soir, c'était une autre nuit, là* ». En troisième lieu, cette rencontre humaine est dite « *profonde* » parce qu'elle incite à un investissement majeur de l'infirmière et de la personne soignée, et ce, à travers l'ensemble du processus de réadaptation de la personne soignée. Le commentaire de Laurence parle de la qualité de la présence de l'infirmière auprès de la personne soignée et de sa famille : « *"Être avec" c'est d'être présente dans tout son ensemble pour le patient, être présente peu importe ce qu'il vit, être présente dans les moments qu'ils vivent présentement. C'est être présente pour le patient, sa famille et l'équipe.* » De même, cette relation humaine profonde, comme expérience de soin infirmier, met en lumière la nature fondamentale de même que complexe de l'expérience d'« être avec » la personne soignée.

Quatrièmement, et toujours selon la perception des participants, cette rencontre se veut *thérapeutique* puisqu'elle répond aux besoins de la personne soignée, optimise la progression de la personne soignée à travers son processus de réadaptation et promeut le niveau de bien-être de la personne (infirmière et personne soignée). Isabelle témoigne de la promotion du bien-être de la personne : « *Pour le patient, le fait d'"être avec" lui l'aide à ne pas se sentir tout seul et* […] *Oui, c'est important de ne pas se sentir seul en situation d'impuissance, si je puis dire, parce qu'à un moment donné, une personne qui croit en toi, déjà, ça fait un monde de différence.* » Mais encore, cette rencontre humaine profonde contribue à l'élévation de la croissance intérieure, c'est-à-dire à la transformation des personnes engagées dans cette expérience de soin. Cette transformation contribue à promouvoir l'harmonie corps-âme-esprit de l'infirmière et de la personne soignée. En définitive, l'expérience d'« être avec » la personne soignée signifie une rencontre humaine profonde et thérapeutique, laquelle s'avère *transformatrice*, non seulement en matière de contribution à la réadaptation de la personne soignée, mais également en fait de croissance intérieure des personnes engagées dans cette expérience de soin exceptionnelle. Le point de vue d'Élaine parle de la métamorphose de l'infirmière : « *Chaque expérience d'"être avec" un patient me permet d'évoluer, de grandir, d'apprendre. Ces acquis me permettent d'aller plus loin avec un autre patient. Ces expériences d'"être avec" mon patient me permettent d'acquérir des habiletés relationnelles.* »

Finalement, cette étude phénoménologique ayant utilisé la méthode Investigation relationnelle *Caring*, elle-même découlant de l'école de pensée de Husserl, offre une description en profondeur de la signification et de la contribution de l'expérience d'« être avec » la personne soignée en réadaptation. L'infirmière, qui travaille dans ce domaine, a accès à une multitude de données probantes afin de guider et de renouveler sa pratique clinique

quotidienne. Par exemple, les résultats éclairent l'infirmière dans le développement et le soutien d'une relation de *caring* entre elle et la personne soignée, dans laquelle s'actualise l'expérience d'« être avec » la personne soignée. Il ressort de cette recherche que c'est uniquement par l'entremise de cette qualité de relation humaine que le soin infirmier contribue à la réadaptation de la personne soignée, de même qu'à la transformation du patient et de l'infirmière.

CONCLUSION

La phénoménologie descriptive et transcendantale de l'école de pensée de Husserl procure un moyen pour décrire et comprendre les questions (p. ex., phénomènes, expériences) qui touchent la signification de l'existence humaine et qui demeurent peu explorées dans la littérature. Elle rend possible l'émergence d'une signification pure et universelle, laquelle illustre le discours commun (structure essentielle) qui se dégage de toutes les entrevues réalisées.

Aussi, la phénoménologie de Husserl comporte certains défis. Par exemple, pour le chercheur, l'attitude de mettre de côté (*bracketing*) ses connaissances, expériences, valeurs et préjugés reliés au phénomène à l'étude reste un défi constant pendant tout le processus de la recherche. De même, tout le processus d'abstraction (réduction eidétique) afin de faire émerger l'essence pure et universelle du phénomène réclame du chercheur une totale immersion dans les témoignages des participants, laquelle demande de prendre le temps nécessaire pour de multiples lectures et relectures des entrevues.

RÉFÉRENCES

BENNER, P. (1994). *Interpretive Phenomenology : Embodiment, Caring, and Ethics in Health and Illness*, Thousand Oaks, Sage Publications.

BROUSSEAU, S., C. CARA et R. BLAIS (2013). « La qualité de vie telle que vécue par des gestionnaires infirmiers de premier niveau œuvrant en centre hospitalier universitaire affilié : une analyse phénoménologique descriptive », communication dans le cadre du 25e Congrès quadriennal du Conseil international des infirmières, Melbourne. Australie.

BUBER, M. (1970). *I and Thou*, New York, Scribners ; *Je et Tu*, Paris, Aubier Montaigne.

CARA, C.M. (1997). *Managers' Subjugation and Empowerment of Caring Practices : A Relational Caring Inquiry with Staff Nurses*, Thèse de doctorat, Boulder, University of Colorado.

CARA, C. (1999). « Relational caring inquiry : Nurses' perspective on how management can promote a caring practice », *International Journal for Human Caring*, vol. 3, nº 1, p. 22-29.

CARA, C. (2002). « Creating a caring environment in nursing research », communication présentée au congrès annuel de l'International Association for Human Caring, Boston, États-Unis.

COLAIZZI, P.F. (1978). « Psychological research as the phenomenologist views it », dans R.S. Valle et M. King (dir.), *Existential-Phenomenological Alternatives for Psychology*, New York, Oxford University Press, p. 48-71.

GADOW, S. (1994). « Whose body ? Whose story ? The question about narrative in women's health care », *Soundings* : *An Interdisciplinary Journal*, vol. 77, nos 3-4, p. 295-307.

GIORGI, A. (1997). « De la méthode phénoménologique utilisée comme mode de recherche qualitative en sciences humaines : théories, pratique et évaluation », dans J. Poupart, L.H. Groulx, J.P. Deslauriers, A. Laperrière, R. Mayer et A.P. Pires (dir.), *La recherche qualitative : enjeux épistémologiques et méthodologiques*, Boucherville, Gaëtan Morin, p. 341-364.

GUBA, E.G. et Y.S. LINCOLN (1981). *Effective Evaluation*, San Francisco, Jossey-Bass.

HUSSERL, E. (1970). *The Crisis of European Sciences and Transcendental Phenomenology*, Evanston, Northwestern University Press.

HUSSERL, E. (1999). « The train of thoughts in the lectures », dans E.C. Polifroni et M. Welch (dir.), *Perspectives on Philosophy of Science in Nursing*, Philadelphie, Lippincott, p. 247-262.

LINCOLN, Y.S. et E.G. GUBA (1985). *Naturalistic Inquiry*, Beverly Hills, Sage Publications.

MUNHALL, P.L. (dir.) (2012). *Nursing Research. A Qualitative Perspective*, 5e éd., Toronto, Jones and Bartlett Learning.

O'REILLY, L. (2007). *La signification de l'expérience d'« être avec » la personne soignée et sa contribution à la réadaptation : la perception d'infirmières*, recherche doctorale, Faculté des sciences infirmières, Université de Montréal, AAT NR 35727.

O'REILLY, L. et C. CARA (2011). « A caring educational intervention with nurses to enhance health care humanization and quality », *International Journal for Human Caring*, vol. 15, nº 3, p. 75.

O'REILLY, L., C. CARA, M.P. AVOINE et S. BROUSSEAU (juin 2010). « Being with » the nurse from the patients' perspective : A phenomenological study in a rehabilitation center applying Watson's caring model », *International Journal for Human Caring*, vol. 14, nº 3, p. 74.

PAILLÉ, P. (2008). « L'examen phénoménologique des données empiriques », dans P. Paillé et A. Muchielli (dir.), *L'analyse qualitative en sciences humaines et sociales*, 2e éd., Paris, Armand Colin, p. 85-98.

PATERSON, J.G. et L.T. ZDERAD (1976). *Humanistic Nursing*, New York, John Wiley & Sons.

RAY, M.A. (1991). « Phenomenological method for nursing research », dans Summer Research Conference Monography, *Nursing Theory Research and Practice*, Detroit, Wayne State University, p. 163-172.

REEDER, F. (1991). « Conceptual foundations of science and key phenomenological concepts », dans Summer research conference monograph, *Nursing Theory Research and Practice*, Detroit, Wayne State University, p. 177-187.

SANDELOWSKI, M. (1986). « The problem of rigor in qualitative research », *Advanced in Nursing Science*, vol. 8, n° 3, p. 27-37.

SPIEGELBERG, H. (1982). *The Phenomenological Movement : A Historical Introduction*, 3e éd., Boston, Martinus Nijhoff.

VAN MANEN, M. (1990). *Researching Lived Experience : Human Science for an Action Sensitive Pedagogy*, Londres, The Althouse Press.

WATSON, J. (1985). *Nursing : Human Science and Human Care*, Norwalk, Appleton-Century-Crofts.

WATSON, J. (1988). *Nursing : Human Science and Human Care*, 2e éd., New York, National League for Nursing.

WATSON, J. (2008). *Nursing : The Philosophy and Science of Caring*, Boulder, University Press of Colorado.

WHITTEMORE, R., S.K. CHASE et C.L. MANDLE (2001). « Validity in qualitative research », *Qualitative Health Research*, vol. 11, n° 4, p. 522-537.

CHAPITRE 3

L'APPROCHE ETHNOGRAPHIQUE
Illustration dans le contexte de la réadaptation en santé mentale

Daniel Côté
Danielle Gratton

FORCES

- Elle permet de décrire en profondeur un phénomène étudié tel qu'il est vécu et perçu au quotidien.
- Elle analyse des facteurs externes qui influencent les émotions, les comportements, les motivations et les réactions individuels.
- Le séjour prolongé sur le terrain permet d'observer de manière fine et complexe une situation donnée et d'assurer la constance de l'analyse.

LIMITES

- La description d'une situation particulière ne permet pas la généralisation.
- La présence sur le terrain demande beaucoup de temps et peut engager des coûts considérables.
- Il y a un risque de se restreindre à un réseau limité d'acteurs et de négliger certains autres (sélectivité).

On peut, avec J.W. Creswell, dénombrer cinq grandes méthodes de recherche qualitatives : les études narratives, la phénoménologie, la théorisation ancrée, les études de cas et l'ethnographie (Creswell, 2007). Certaines de ces approches sont décrites ailleurs dans cet ouvrage collectif. Nous nous pencherons ici sur l'ethnographie, qui nous semble moins connue dans le champ de la réadaptation et en santé mentale. Nous tenterons de définir l'ethnographie, d'en dessiner les contours et de mettre en lumière le fonctionnement de cette approche méthodologique.

L'ethnographie s'est développée historiquement en anthropologie (ou ethnologie). Cette dernière visait à mieux comprendre les modes de vie des sociétés non occidentales ou dites « traditionnelles ». L'ethnographie a été l'outil de recherche privilégié pour y parvenir puisqu'elle permet d'observer *in situ* et en prenant part à la vie quotidienne des peuples qu'elle tente de comprendre. Comme on le verra, l'ethnographie n'élimine pas la pertinence de l'étude des facteurs idiosyncrasiques, c'est-à-dire des facteurs qui sont propres à l'individu (affects, émotions, perceptions), mais elle offre une approche complémentaire qui permet d'explorer les logiques qui fondent une société et les sens partagés (Bateson, 1972), et les rapports entre le psychisme, la biologie, la société et la culture (Massé, 1995). Cette complémentarité est très utile, notamment quand il s'agit de comprendre les facteurs qui influencent la réinsertion sociale.

L'ethnographie est une discipline qui cherche à comprendre un univers complexe et dont les approches visent à cerner les différences entre les êtres humains (Agar, 1982). Pour repérer les caractéristiques des constructions sociales, elle a développé une variété de méthodes, de techniques d'enquête et d'analyse. Elle produit et analyse des données descriptives à partir de la parole échangée lors de contacts de proximité, ou bien à partir de l'observation directe de situations d'interactions, ou encore celle de contextes de vie quotidienne. Pour faire sa collecte de données, l'anthropologue note dans un journal de bord, par exemple, ses observations, qui lui servent à mettre en évidence l'imbrication complexe entre les catégories de l'expérience personnelle et les catégories culturelles, c'est-à-dire partagées dans des communautés territoriales (pays, quartiers, institutions, etc.), ou encore d'identité (jeunes, itinérants, personnes souffrant de maladie mentale, etc.), faisant apparaître des milieux urbains complexes qui, souvent, servent de points d'ancrage et de structuration à ces expériences. L'ethnographie tente le plus souvent possible de décrire ou de reconstituer les modes de vie et l'univers de sens et de pratiques (pensées, croyances, représentations) des personnes dans les termes et les expressions qui leur sont propres, tout en

situant cet univers dans un champ de pratiques et de représentations sociales (environnement social et culturel, milieu professionnel, etc.) (Symon et Cassell, 1998).

Ce chapitre se divise en deux sections : la première aborde des questions d'ordre épistémologique et paradigmatique ; elle présente les fondements théoriques de la recherche ethnographique, les méthodes de collecte de données, et les exigences pratiques d'un devis ethnographique et ses critères de scientificité. La deuxième section porte sur l'application de l'approche ethnographique pour répondre à des questions de recherche bien précises. Pour y parvenir, nous décrirons et analyserons une étude de réadaptation en santé mentale qui se réclame d'une approche ethnographique. L'objectif de cette étude sera présenté ainsi que le contexte dans lequel elle a été conçue, de même que les instruments de collecte et d'analyse des données. L'apport de l'approche ethnographique sera mis en évidence. En parcourant cette section, le lecteur pourra en concevoir les variantes et comprendre les forces et les limites de cette approche par rapport aux autres méthodes de recherche qualitative.

1. ORIGINE ET FONDEMENTS THÉORIQUES DE L'APPROCHE ETHNOGRAPHIQUE

L'ethnographie s'est développée en anthropologie à la fin du XIXe siècle et est devenue la pièce maîtresse de cette discipline tout au long du XXe siècle. Pour comprendre des modes de vie différents, les anthropologues font de longs séjours sur le « terrain » à étudier les us et coutumes des peuples avec lesquels ils entrent en contact, en focalisant parfois sur des phénomènes précis comme la musique, la religion populaire, les structures de la parenté et de l'organisation familiale, la mythologie, mais aussi la médecine, la santé et la guérison (Lenclud, 2000). L'ethnopsychiatrie est une des spécialités anthropologiques ; elle constitue un champ de recherche consacré à l'étude de la santé mentale qui repose largement sur l'utilisation de données ethnographiques. Elle se consacre notamment à l'étude des phénomènes liés à la « folie », à la transe, au chamanisme, à la possession, ainsi qu'aux mécanismes psychologiques et sociaux qui les sous-tendent (Deluz, 2000)[1]. Au Québec, Corin (Corin, 1990 ; Corin et Lauzon, 1994) s'est particulièrement intéressée au vécu des personnes souffrant de schizophrénie et aux exigences que représente l'autonomie pour ces personnes.

1. En Occident, l'ethnopsychiatrie a contribué au développement de cliniques de psychiatrie transculturelles qui permettent de lire autrement l'expérience de la souffrance en accordant, par exemple, au délire psychotique une attention particulière. Certaines écoles de pensée, lacaniennes notamment, vont analyser un délire psychotique à la manière d'un mythe ou d'un récit pour en reconstituer le sens caché ou pour amener l'utilisateur de services à verbaliser le malaise qui l'afflige.

L'ethnographie est appelée aussi « observation participante », car de façon traditionnelle, dans cette discipline, il s'agit de vivre dans des groupes humains pour mieux saisir leurs caractéristiques propres. Cette méthode a été employée par des anthropologues comme Gregory Bateson (1972) et Erving Goffman (1961). Goffman s'est intéressé aux marques qui engendrent l'exclusion et la stigmatisation. Il est d'ailleurs connu pour avoir séjourné dans une unité de psychiatrie, dans les années 1950, afin d'en comprendre le fonctionnement. Son livre *Asylums* (1961) fait une description ethnographique de ce milieu. Un film américain, *One Flew over the Cuckoo's Nest* (*Vol au-dessus d'un nid de coucou*, 1975) découle également de ses travaux. L'apport de cet anthropologue réside dans l'humanisation des services en santé mentale. Comme Bateson, il a collaboré, notamment avec des psychiatres et des psychologues à l'École de Palo Alto, au développement d'une nouvelle théorie de la communication et des relations interpersonnelles[2]. De nos jours, l'ethnographie n'est plus une chasse gardée de l'anthropologie sociale et culturelle ni des autres sciences sociales ; diverses sciences de la santé tentent de s'approprier ses principes et ses exigences pratiques afin de mieux cerner les modes de vie des personnes dites vulnérables ou qualifiées parfois de « marginales », de même que les exigences sociales à leur égard dans les milieux urbains.

Plusieurs paradigmes ou modèles épistémologiques peuvent orienter l'approche ethnographique. Des paradigmes comme le positivisme, la théorie critique ou le constructivisme, pour ne citer que ceux-là, peuvent influencer la manière de construire l'objet de recherche et le choix des outils d'enquête. Depuis la deuxième moitié du XXe siècle, le constructivisme a grandement influencé les chercheurs en sciences sociales, qui considèrent que la connaissance n'est jamais purement neutre et objective et qu'elle ne peut pas faire abstraction de la sensibilité et de la subjectivité du chercheur et de l'ancrage de ce dernier dans un environnement social, culturel, politique et économique donné (Burr, 2003). Cet environnement possède des structures, des normes et des modèles de fonctionnement qui se sont développés au fil du temps et qui demeurent en mouvement constant. La science, comme système et comme construction historique, n'y échappe pas. Le paradigme constructiviste accorde aussi à l'individu la capacité d'appréhender la réalité et de la transformer.

2. C'est grâce à cette forme de recherche qualitative que l'approche systémique, bien connue des intervenants et des thérapeutes en santé mentale, a été développée. Le concept de « double contrainte » (*double bind*) vient aussi de l'observation participante, car c'est Bateson (1972), un anthropologue, qui a observé, pendant qu'il était en interaction avec des familles « schizophrènes », un mode de communication sans issue, typique de ce genre de familles. Ce concept est lui aussi bien connu dans le milieu de l'intervention familiale et en santé mentale.

Dans ce paradigme, l'action de l'intervenant se déplace de l'individu aux situations à changer. L'objectif de la recherche ethnographique vise à comprendre le sens que des individus, dans une même situation, un même groupe, donnent aux événements et aux situations de leur vie quotidienne ; la recherche tente de comprendre par quels processus et dans quelles circonstances se construit le sens ou la représentation d'une situation donnée, comme toute l'expérience émotionnelle qui s'en dégage (Chwalisz, Shah et Hand, 2008). Car, pour reprendre les termes du sociologue américain Herbert Blumer (1969), c'est en fonction du sens donné aux choses qui les entourent que les gens orientent leurs conduites. La théorie de la stigmatisation en santé mentale découle de ce même paradigme (Goffman, 1975). L'anthropologie s'inspire généralement de ce paradigme, en admettant que le sujet, s'il est influencé par le contexte dans lequel il vit, peut aussi y jouer un rôle actif et structurant ; ce type d'études comporte souvent une visée émancipatrice (*empowerment*) en faisant connaître des situations qui étaient restées dans l'ombre jusque-là, et cela, même dans le contexte de la rencontre clinique, comme c'est le cas avec Kleinman (1980)[3]. Certains auteurs ont développé une approche ethnographique dite « institutionnelle » qui cherche à illustrer le point de vue et le vécu d'une pluralité d'acteurs qui se situent les uns par rapport aux autres dans des positions hiérarchiques ou dans des rôles différents (thérapeutes, gestionnaires, utilisateurs de services, etc.) (Moll *et al.*, 2012 ; Smith, 2002). Une approche ethnographique peut dès lors décrire la manière par laquelle une certaine configuration des rapports sociaux se met en place, derrière le jeu complexe des interactions interpersonnelles (Smith, 2005).

1.1. Pourquoi choisir l'ethnographie parmi les approches qualitatives ?

En général, la méthode ethnographique est utilisée quand une situation nous échappe. À la différence de la phénoménologie (voir le chapitre 2 de cet ouvrage), plus centrée sur le vécu subjectif des personnes et sur la dimension narrative de l'expérience, l'approche ethnographique permet de cerner des contextes plus larges et des situations données où les acteurs prennent place et interagissent ; l'unité d'analyse est davantage une collectivité donnée

3. Le psychiatre et anthropologue américain Arthur Kleinman a développé le concept de « modèles explicatifs de la maladie » (*illness explanatory model*) à la fin des années 1970 (Kleinman, 1980). Ce concept comprend neuf rubriques, dont la perception des causes du problème, la manière de le nommer, l'origine présumée de ce problème, ses effets sur la personne souffrante, les peurs qui lui sont associées, la perception du traitement idéal, les attentes de résultats par rapport au traitement actuellement reçu, etc. Ce concept a été repris dans l'élaboration d'une grille d'entrevue par des chercheurs montréalais en psychiatrie transculturelle (Groleau, Young et Kirmayer, 2013). Cette grille remet en question explicitement la relation thérapeute-patient, la dimension affective liée au problème, la démarche de recherche de solutions et tout l'univers social dans lequel la maladie prend place et se définit.

que l'individu proprement dit, même s'il faut passer par la collecte des perceptions individuelles pour y parvenir (Spencer, Krefting et Mattingly, 1993). Elle propose donc un regard plus systémique sur une réalité donnée (Mucchielli, 1998).

Dans le champ très vaste de la réadaptation, physique ou mentale, et comme dans les autres disciplines de la santé, il est reconnu que l'attitude des utilisateurs de services, leurs représentations de la maladie, leur motivation à suivre un traitement et leurs attentes de résultats par rapport au traitement reçu influencent grandement le processus de rétablissement (Coutu *et al.*, 2007). Or, ceci peut être particulièrement problématique dans des relations thérapeutiques où clients et thérapeutes ne partagent pas les mêmes référents culturels (Sloots *et al.*, 2010; Pooremamali, Persson et Eklund, 2011). Une recherche ethnographique peut alors s'intéresser aux caractéristiques de la rencontre thérapeutique et en définir les contours (formes de communication verbales et non verbales, stratégies mises en œuvre des programmes, contraintes institutionnelles, etc.) pour porter la réflexion à la fois sur l'adhésion au traitement au-delà de la désignation de facteurs strictement personnels, et sur le plan des interactions interpersonnelles et de possibles contraintes systémiques ou organisationnelles (MacEachen *et al.*, 2010).

1.2. Exigences pratiques d'un devis ethnographique et quelques étapes à suivre

Faire de l'ethnographie, c'est d'abord choisir un terrain d'enquête à partir duquel et à travers lequel prendront place des événements plus ou moins familiers, parfois même étrangers, où l'ethnographe sera placé en situation d'interaction constante avec des personnes qui deviendront ses informateurs clés (Winkin, 2001). Toutefois, l'ethnographe se heurte souvent à un enjeu méthodologique de taille, qui est d'obtenir l'accès au site et de définir son rôle sur le terrain (observateur, intervenant, etc.) (Silverman, 2011). C'est que l'ethnographe est rarement attendu sur le terrain. Et malgré les autorisations officielles qu'il a pu obtenir préalablement, sa présence peut être gênante pour certaines personnes, qui se sentent observées, notées, évaluées, ou épiées. De plus, il peut être difficile pour les participants à ce type d'études de comprendre ce modèle de recherche. C'est pourquoi il est souvent préférable d'attendre quelques semaines avant de solliciter des entrevues et de sortir carnets de notes, enregistreurs et autres appareils audiovisuels. L'ethnographe doit gagner la confiance des personnes du milieu à l'étude; il doit même, idéalement, arriver à se faire oublier.

L'ethnographie comporte certaines exigences pratiques. L'ethnographie exige une présence sur le « terrain » (clinique, milieu de travail, centre de jour, etc.) (Nader, 2012) pour observer *in situ* ce qui s'y passe, ce qui se dit, qui fait quoi, dans quelles circonstances et à quel moment précis. L'ethnographie devient ainsi indissociable de l'« observation participante » à laquelle elle est souvent associée, voire assimilée (Laperrière, 1995 ; Mayer et Ouellet, 1991), ce qui peut sembler en opposition avec des formes de recherche qui se veulent plus « objectives ». L'ethnographie est donc une forme d'immersion du chercheur dans la situation sociale qu'il étudie (Laperrière, 1995). Souvent décrite comme une « théorie de la description » (*theory of description*), l'ethnographie doit rendre compte d'une situation dans ses plus menus détails pour en exposer les dynamiques et les modes de fonctionnement (*patterns*), les récurrences et les contradictions. Car on passe déjà à un autre niveau de construction de la réalité vécue.

L'ethnographie ne se réduit toutefois pas à la seule observation participante. Elle s'appuie en cela sur l'utilisation de plusieurs outils de collecte des données : observation directe, observation participante (qui implique de jouer un rôle social ou d'accomplir une tâche pendant l'observation, comme accompagner une personne à une visite médicale pour observer les interactions), entretiens semi-dirigés, dialogues et échanges informels, entretiens en profondeur, récits de vie, collecte de matériaux produits par les personnes sur le terrain (p. ex., audiovisuel, documents d'information, etc.), photographies, vidéos (Silverman, 2011). Cette diversité d'outils de collecte de données permet à l'ethnographie de varier ses sources d'informations et d'assurer ainsi la triangulation de ses résultats (exemples et contre-exemples).

De nos jours, même si l'exigence du terrain est discutée au sein même de l'anthropologie, la relation entre l'ethnographe et l'informateur demeure fondamentale. Une bonne ethnographie repose sur cette capacité à rendre compte ou à refléter l'univers vécu par ces personnes, dans leurs propres mots ou dans un langage qui traduit bien ce qu'ils ressentent et perçoivent par rapport à une situation donnée. C'est ainsi que de nombreux anthropologues qui se réclament de la méthode ethnographique utilisent l'entretien en profondeur comme principale source de données. Dans l'entrevue, le sujet ne vit pas la situation que l'on cherche à décrire et à comprendre, il est déjà dans un moment analytique, dans une sorte d'anamnèse, avec toutes les limites que cela comporte en matière de mémoire, de désirabilité sociale ou d'autocensure. Dans ce contexte, l'entretien peut être qualifié d'ethnographique, comme c'est le cas de l'étude présentée dans la section 2 de ce chapitre. L'entretien en profondeur offre une vitrine riche et précieuse sur l'univers de sens des personnes interrogées et la manière dont ces dernières

se situent dans un environnement social donné. Ces entretiens se déroulent généralement dans le contexte de vie des personnes interrogées, soit à leur domicile, au travail, au parc de quartier, dans un centre de service local (clinique externe, centre de jour en santé mentale, centre d'emploi, etc.) ou dans un commerce local (épicerie, bar, café, etc.). Le contexte de déroulement de l'entrevue doit être décrit en détail; l'ethnographe ne doit pas négliger non plus d'y étaler ses impressions et ses sentiments personnels. C'est à ce moment qu'entrent en scène la prise de notes et l'écriture ethnographique.

1.3. Prise de notes et écriture ethnographique

En quoi consiste alors l'écriture ethnographique? L'écriture ethnographique repose d'abord sur la prise de notes (*fieldnotes*), sur la tenue d'un journal de bord et sur l'interprétation des observations recueillies (Spradley, 1980). Le journal de bord est un outil précieux, car il permet d'explorer de nouvelles situations sans préconceptions et de s'assurer de garder des données dont la valeur ne peut souvent être connue qu'à la fin de l'entreprise de recherche.

L'entrée sur le terrain ou l'immersion dans l'univers des personnes interrogées est toujours déstabilisante pour un ethnographe, qui peut éprouver beaucoup d'incertitudes et d'incompréhension. Au moment d'entreprendre une étude de type ethnographique, la prise de notes est désordonnée; l'ethnographe note tout ce qui se passe. Lorsque l'ethnographe consigne une observation dans un journal de bord, il doit s'assurer que chacune de ses observations est suffisamment décrite, soit en précisant le lieu de la collecte, le contexte, le moment et l'identité des personnes impliquées (leur rôle, leur statut, etc.) (Laperrière, 1995; Mayer et Ouellet, 1991). Il note également comment se passe son intégration dans le milieu, ses expériences diverses, ses impressions personnelles, ses sentiments (peurs, angoisses, etc.); et il note comment réagissent les personnes qu'il rencontre par rapport à une situation donnée (p. ex., recherche d'emploi, expérience du stigmate social), comment ils structurent leur vie quotidienne et à partir de quels préceptes, valeurs ou idéaux (Mayer et Ouellet, 1991). Le journal de bord constitue selon Emerson, Fretz et Shaw (2011) le premier instrument du chercheur pour l'aider à comprendre et à analyser une situation observée; plus encore, c'est le moyen de consigner l'évolution de cette analyse et comment on parvient à construire et à tester des hypothèses de recherche (Emerson *et al.*, 2011). Mayer et Ouellet (1991) distinguent en cela des notes personnelles, des notes descriptives et des notes théoriques. Les notes personnelles portent davantage sur l'expérience émotionnelle du chercheur, qui y exprime ses angoisses, ses peurs et ses frustrations, ainsi que ses opinions; ces notes sont chargées émotionnellement et sont porteuses

de biais qui peuvent influencer le déroulement de la recherche (Spradley, 1980); c'est pourquoi elles doivent être rapidement revues et corrigées en les exposant de façon répétée au terrain.

Les notes descriptives sont un compte rendu de rencontres, de lieux et des sites observés, de conditions particulières dans lesquelles des rencontres et des interactions ont eu lieu; elles doivent être les plus détaillées et les plus neutres possibles. Chaque situation observée (événement, rencontre, entrevue, etc.) doit être datée et située clairement dans son contexte pour qu'il soit aisé de relire, de comprendre et de situer le contexte de la prise de notes plusieurs mois plus tard, voire des années (p. ex., quelles personnes sont présentes lors de l'événement que l'on décrit, quel est leur statut, quels sont les liens hiérarchiques entre elles, qui fait quoi et en quelles circonstances, de quelle manière s'articulent les différences de genre, les rapports entre générations, le langage non verbal, l'expression des émotions, les gestuelles, etc.). Au fil du travail sur le terrain, les observations deviennent plus sélectives et systématiques (Sanjek, 1996c); l'ethnographe commence à comprendre le milieu qu'il étudie et les schémas cognitifs, corporels, comportementaux, relationnels ou émotionnels qui y régissent la vie quotidienne. Des hypothèses prennent forme. Ce sont les notes théoriques, c'est-à-dire des notes analytiques et interprétatives où le chercheur commence à émettre une vision plus générale sur le phénomène étudié et sur la signification qu'il revêt auprès des personnes qu'il côtoie et qu'il commence à connaître, parfois même intimement (Emerson *et al.*, 2011).

Il existe diverses façons de rédiger des notes de terrain et d'organiser le matériel recueilli. Certains chercheurs peuvent commencer leur travail de description ethnographique en rapportant les détails d'une routine quotidienne (Sanjek, 1996b). Plusieurs éléments de la vie quotidienne peuvent constituer une manière d'indexer les notes prises par le chercheur (p. ex., l'environnement de travail, les activités de loisirs, les conflits). Il n'est pas nécessaire d'attendre la fin du terrain pour entamer l'indexation et la codification du matériel; au contraire, cette étape se fait de manière simultanée avec la période de collecte des données et permet d'aller vérifier sur le terrain des observations et les premières hypothèses. Cette démarche, dite « itérative », constitue un canon de la méthode de la théorisation ancrée (*grounded theory*) (Annells, 1996), qui s'est appuyée à ses débuts sur la pratique du terrain et de l'observation participante.

1.4. Critères de scientificité

Toute recherche scientifique repose sur des critères de scientificité pour asseoir sa rigueur et sa valeur heuristique. Il existe dans toutes les sciences des débats sur la valeur et la pertinence des approches existantes. En psychologie, par exemple, on oppose les approches psychodynamiques aux approches humanistes et cognitives-comportementales, ou encore aux approches systémiques. En anthropologie, le débat oppose principalement les approches modernistes et postmodernistes. La notion de culture est au centre de ces débats (Cuche, 2001). Les postmodernistes reprochent, par exemple, aux modernistes de donner une impression figée de la culture[4]. En fait, il faut reconnaître, comme en psychologie, que chaque approche a ses forces et ses limites. C'est la position défendue par l'approche herméneutique, qui préconise l'usage de plusieurs modèles complémentaires pour cerner un phénomène (Nader, 2012; White, 2012).

Les penseurs postmodernistes soutiennent que la réalité ne peut être appréhendée objectivement puisque la description et l'analyse de cette dernière reposent en grande partie sur les postulats théoriques du chercheur (p. ex., fonctionnalisme, constructivisme, phénoménologie, postmodernisme) et constituent autant de points de repère dans l'interprétation des phénomènes sociaux (Nader, 2012). Malgré ces divergences, être capable de positionner un devis dans un cadre théorique plus général est reconnu comme un critère de rigueur, car cela permet au chercheur de dresser ses propres horizons et d'en reconnaître les limites (Malterud, 2001).

Les anthropologues jouissent généralement d'un solide bagage épistémologique qui les rend conscients de leurs propres orientations théoriques et capables d'une réflexion critique très pointue. Sensibilisés à cet aspect par les postmodernistes, ils ne craignent pas de faire leur examen autocritique lorsqu'ils présentent leurs données. Cette prise de conscience de leurs propres biais par les chercheurs peut paraître évidente du côté des sciences de la santé, car, en général, les intervenants et les professionnels reconnaissent les limites de leurs modèles théoriques. Mais ils peuvent être moins habitués à s'interroger sur les fondements de leur culture sociale et professionnelle. Par exemple, peu d'entre eux s'interrogent sur la notion d'autonomie, qui oriente la réadaptation et qui trouve son sens dans la société occidentale (Gravel *et al.*, 2009). Étant particulière à notre type de société, cette notion demande à être revue avec des clients de différentes

4. Que l'on peut définir comme des façons différentes de penser, de dire et de faire qui se sont construites historiquement et se transforment continuellement sous la poussée de contingences et de jeux de pouvoir.

origines ethnoculturelles, car elle risque de poser des problèmes éthiques quand une personne s'est construite sur un autre mode relationnel (p. ex., interdépendance) (Iwama, 2003).

La science est aussi un discours sur les conditions de la production du savoir, pas juste un étalage de techniques et de procédés éprouvés qui peuvent donner une impression de neutralité et d'objectivité (Courtois et Desmet, 2007). L'approche ethnographique a été abondamment critiquée en anthropologie, tantôt comparée à un genre littéraire, tantôt assimilée à une œuvre, rigoureuse certes, mais de fiction (Geertz, 1995 ; Marcus et Fischer, 1986 ; Clifford, 1986 ; Strathern, 1990 ; Narayan, 2008). Fait ou fiction, l'intention ethnographique demeure noble : décrire et comprendre un phénomène dans les termes de ses propres commettants pour en extraire son sens et sa logique propre.

Winkin (2001) nous rappelle que l'ethnographie est une expérience personnelle qui procède d'une éducation du regard et de l'écriture ; ce couplage regard/écriture va de pair avec la nécessaire retraduction de cette expérience pour un public tiers qui n'a pas vu, qui ne connaît pas et qui possède peu de moyens pour attester ces observations tant elles sont ancrées dans l'expérience prolongée du chercheur sur le terrain. La crédibilité des observations et du matériel ethnographique dans son ensemble (données d'entrevues, vidéos, etc.) repose, comme le rapporte Sanjek (1996), sur la plausibilité du récit et sur son caractère convaincant, sur sa clarté et sur sa cohérence (absence de contradictions internes et illogismes). La rhétorique l'emporte-t-elle sur la validité ? Pas nécessairement. Les conditions de la recherche ethnographique sont certes difficilement reproductibles et falsifiables au sens poppérien (Popper, 1959), mais elles reposent sur des exigences relationnelles, affectives, et sur la capacité du chercheur à faire correspondre les résultats de ses propres observations aux théories explicatives existantes et de montrer le lien, continu ou discontinu, entre la situation observée et ce qui a déjà été décrit dans la littérature. C'est le sens de la démarche anthropologique, avec ses limites et ses écueils épistémologiques.

Avec l'ethnographie, le critère de « transférabilité » ou la capacité de généraliser est difficile à appliquer puisque l'échantillonnage n'est pas représentatif de l'ensemble de la population. Malgré cela, l'approche ethnographique doit tout de même se soucier de la fiabilité des observations qui sont faites sur le terrain et s'assurer qu'elles reflètent la réalité que l'on tente d'observer, dont elle vise à cerner les contours. Elle doit s'assurer de décrire en profondeur le phénomène étudié et que la collecte des données a atteint un point de saturation (lorsque la collecte d'information n'apporte plus de nouveaux résultats ni de nouvelles analyses). Pour répondre aux exigences de scientificité, McReynolds *et al.* (2001) proposent neuf critères :

1. traçabilité des données (accès aux données brutes);
2. notes de terrain (description détaillée des situations observées) (pour plus d'informations sur la prise de notes, voir le recueil de Sanjek, 1996a);
3. rédaction de mémos (réflexions sur les biais, les pistes d'interprétation et les impressions du chercheur);
4. triangulation (la même information est obtenue de sources différentes et variées);
5. recours à plusieurs chercheurs (chaque chercheur apporte sa propre expertise et un regard qui lui est propre sur l'interprétation des résultats);
6. recours à de multiples sources de données (entrevues, observation, documents audiovisuels, etc.);
7. recherche de données contraires (certaines données soutiennent ou discréditent une hypothèse de recherche, les données contraires aident à raffiner et à modifier une théorie émergente);
8. vérification des résultats auprès des participants (les participants se reconnaissent-ils dans les résultats présentés? C'est également une occasion d'apporter des précisions aux données recueillies); et
9. discussion entre pairs qui ne font pas partie du projet (autres perspectives, distance critique par rapport au terrain).

L'ethnographie devient une ressource importante au moment, où, par exemple, les sociétés changent. Ou encore, lorsque la diversité culturelle devient une caractéristique des clientèles visées par les programmes de santé et de services sociaux et que des valeurs ou des normes antagoniques (p. ex., autonomie et interdépendance) peuvent poser un obstacle au processus thérapeutique (Côté, 2013; Gratton, 2009).

2. ÉTUDE DE RÉADAPTATION EN SANTÉ MENTALE

La prochaine section présente une étude de réadaptation en santé mentale réalisée dans la région des Grands Lacs par des anthropologues américains (Jenkins et Carpenter-Song, 2008). Ils étudient comment des personnes ayant une schizophrénie mettent au point des stratégies reconnues de protection et de résistance dans leurs rapports interpersonnels. Les auteurs étudient ces stratégies à partir d'une approche intersubjective et ils s'intéressent aux circonstances qui déclenchent ces mécanismes, tout comme à

leurs conséquences émotionnelles et personnelles. Cette étude est certes différente des monographies anthropologiques classiques, car elle est adaptée pour le milieu institutionnel et urbain. Elle se distingue des études qualitatives que l'on retrouve le plus souvent dans le champ de la réadaptation et c'est pourquoi elle offre un regard neuf sur des situations courantes qui, autrement, passeraient inaperçues (Hammersley et Atkinson, 2007).

Jenkins et Carpenter-Song (2008) explorent l'expérience vécue par des personnes aux prises avec une schizophrénie ou des troubles schizoaffectifs et mettent l'accent sur la stigmatisation (Goffman, 1975). L'étude part du constat de la persistance de l'exclusion malgré une deuxième génération de médicaments antipsychotiques qui contrôle mieux les symptômes de cette maladie, comme la clozapine et le rispéridone. Les auteurs de l'étude veulent dépasser et éviter l'accent qui est souvent mis en psychologie sur les attributs personnels des personnes ayant des symptômes typiques de cette maladie pour expliquer leurs difficultés d'insertion. Pour ce faire, les auteurs se penchent plutôt sur les expériences de vie quotidienne et ils centrent leurs observations sur des situations réelles d'interaction. Les chercheurs utilisent trois méthodes de collecte de données. Dans un premier temps, le SEMI (Subjective Experience of Medication Interview), un schéma d'entrevue de type semi-ouvert, sert à effectuer le choix des personnes qui peuvent participer à cette recherche. Pour des fins scientifiques, elles doivent être reconnues comme des personnes ayant reçu un diagnostic de schizophrénie selon les critères du DSM-IV (APA, 2000), et leurs symptômes doivent être contrôlés. Dans un deuxième temps, les auteurs ont recours à des entrevues en profondeur pour mettre en évidence des perceptions ou des situations vécues d'exclusion qui mettraient en relief le phénomène de la stigmatisation.

La sélection des participants se fait donc en deux temps : les auteurs s'intéressent aux expériences entourant la prise de médicament et le traitement, soit les situations vécues, les activités quotidiennes, la gestion de la maladie et les relations sociales. Ils les mettent en rapport avec plusieurs données sociodémographiques comme le genre et l'identité. Les entretiens se déroulent dans le cadre de deux centres communautaires en santé mentale qui œuvrent auprès d'une clientèle multiethnique. Ils sélectionnent ainsi 49 hommes et 41 femmes, dont 70 Euros-Américains et 20 Afro-Américains. Dans un deuxième temps, avec des entretiens dits ethnographiques, ces chercheurs s'intéressent précisément au vécu entourant les stigmates sociaux. Ils tentent de savoir avec chaque personne rencontrée si elle vit des stigmates sociaux. Pour ces auteurs, on est en présence d'un phénomène de stigmatisation lorsqu'une des conditions suivantes est présente : le sujet indique que, malgré un contrôle des symptômes de sa maladie, des personnes de son

entourage réagissent négativement à sa maladie; le sujet indique qu'il sent le besoin de cacher sa maladie ou le fait qu'il prend des médicaments pour contrôler les symptômes de sa maladie; le sujet exprime clairement qu'il a perçu des réactions négatives en lien avec sa maladie. Quatre-vingt-seize pour cent des personnes rencontrées rapportent des situations qui attestent du phénomène de la stigmatisation et de l'exclusion. Dans ces entretiens en profondeur, les attentes concernant la récupération et la qualité de vie sont aussi explorées. Voici quelques exemples de questions qui peuvent être posées en employant le SEMI: «Avez-vous l'impression d'avoir le contrôle sur votre maladie?» «Que faites-vous pour vous aider à contrôler votre maladie?» «Comment vos amis réagissent-ils aux problèmes qui sont associés à votre maladie?» «En plus d'avoir à vous soucier de votre propre état/maladie, devez-vous prendre soin de quelqu'un d'autre en ce moment?» (Jenkins et Carpenter-Song, 2005). Des questions de relance et de précisons sont prévues lorsque les réponses sont trop courtes ou imprécises. Le SEMI permet d'obtenir des données narratives sur des sujets variés relatifs à l'expérience de la maladie: expérience de la médication, situation de vie immédiate, activités quotidiennes, relations sociales, gestion de la maladie, attentes de rétablissement, stigmates sociaux, etc. Les participants ont été interrogés jusqu'à trois reprises, pour une durée d'entrevue variant entre une heure et demie et deux heures. Enfin, une observation sur le terrain qualifiée de «naturaliste» vient augmenter les données recueillies avec les deux méthodes précédentes. L'observation dite «naturaliste» consiste à accompagner les participants dans leurs activités quotidiennes et rendez-vous ponctuels (p. ex., supermarché, église, buanderie, restaurant, clinique). Les entrevues ont été menées dans ce contexte de vie quotidienne. Les notes de terrain[5] et les données d'entrevues ont été transcrites et analysées en suivant la méthode de la «théorisation ancrée» (*grounded theory*) (voir le chapitre 5 de cet ouvrage).

L'étude confirme l'utilisation de toute une série de stratégies pour se protéger contre la stigmatisation: dissimuler le diagnostic et la prise de médicaments; faire de l'évitement social; adopter une attitude qui vise à «passer pour normal»; diminuer l'importance de la maladie ou la relativiser; faire de l'enseignement aux personnes de l'entourage; socialiser avec des personnes ayant la même maladie ou avec des personnes empathiques; à l'inverse, affronter ou s'opposer; ou employer le ton de l'autodérision et faire des farces; enfin, reproduire le stigmate social envers d'autres

5. Les données rapportées par Jenkins et Carpenter-Song viennent surtout des entretiens en profondeur; les données d'observation et la prise de notes paraissent mal dans leur étude et laissent supposer, à tort, qu'ils ont été des observateurs neutres, imperturbables, et qu'ils n'ont pas été «absorbés» ou «déstabilisés» par leur travail de terrain. D'autres études vont montrer davantage la subjectivité du chercheur en situation d'observation sur le terrain (Babiss, 2002).

personnes ayant une maladie mentale. Cette dernière stratégie permettrait de comparer la perception subjective de soi aux comportements des autres ayant la même maladie et de souligner son propre rétablissement. Cette stratégie de différenciation (*othering*) implique, selon Jenkins et Carpenter-Song (2008, p. 390-391), une mise à distance de l'étiquette de « malades mentaux » que l'on réserve à ceux que l'on juge plus mal en point[6]. Les données recueillies confirment que les stratégies liées aux stigmates sont intersubjectives, et dépendent de la perception populaire par rapport à la schizophrénie et de l'interaction intersubjective entre les personnes qui présentent une schizophrénie et celles de l'entourage, comme le met en évidence Christy, une participante euro-américaine âgée de 26 ans :

> *Je pense que les gens comprennent le diabète. Ils comprennent aussi le cancer. Mais quand ils entendent parler de maladie mentale, ils deviennent, pas tous, mais ils ne voient pas ça comme le cancer ou le diabète. Vous voyez un psychiatre[7] !* (Jenkins et Carpenter-Song, 2008, p. 393)

Cette perception populaire influence le regard que portent les personnes malades sur elles-mêmes, comme c'est le cas pour Alicia, une Afro-Américaine de 33 ans, qui se voit comme « *paresseuse ou irresponsable* » quand elle n'arrive même pas à se rendre aussi souvent que nécessaire à ses rendez-vous médicaux. Alicia, qui est croyante, se demande aussi si le fait d'être malade ne fait pas d'elle une mauvaise chrétienne.

Selon les auteurs, le cas de Steven, un Euro-Américain de 42 ans, met en évidence comment la schizophrénie remet en question des attentes en matière d'autonomie et de rôles sociaux au prix du déni de la maladie par les membres de la famille :

> *Mes frères et sœurs n'ont pas gardé de contact avec moi. Ils n'écrivent pas, et ne téléphonent pas. Ils ne me rendent pas visite, non plus. J'ai été 19 fois à l'hôpital et pas une seule fois, ils ne sont venus me rendre visite. Et cela fait beaucoup réagir les gens quand je leur raconte que ma famille s'en fout. Ils s'attendent à ce que je me comporte comme si je n'avais pas de schizophrénie… Ils ne veulent jamais en parler ou en discuter. Ils n'abordent jamais le sujet. C'est comme s'ils voulaient que je sois normal, même si je suis malade.* (Jenkins et Carpenter-Song, 2008, p. 396)

D'autres personnes avec une schizophrénie rapportent que plusieurs leur disent de cesser de prendre des médicaments, car ils ne comprennent pas leur importance pour le contrôle de leurs symptômes. Avec ce type de données, les auteurs concluent que les facteurs de réussite personnelle et de réinsertion professionnelle et sociale ne dépendent pas seulement de

6. Cette stratégie possède un versant opposé, qui consiste à s'autostigmatiser et à nourrir une image négative de soi.
7. Tous les témoignages tirés de l'étude de Jenkins et Carpenter-Song sont des traductions libres.

facteurs médicaux ou psychiatriques, mais aussi du fait que les préjugés envers les personnes ayant une maladie mentale persistent, ce qui contribue à faire obstacle à la réinsertion. Les auteurs rapportent que le phénomène de l'anticipation du rejet peut contribuer également à faire obstacle à la réinsertion. L'étude de Jenkins et Carpenter-Song (2008) valide des données qui vont à l'encontre de théories généralement acceptées, en mettant en évidence à partir des expériences rapportées par les participants et de leurs observations, et à l'instar des travaux de Corin (1990) au Québec, comment l'autonomie personnelle, une valeur profondément ancrée en Occident et plus particulièrement en Amérique du Nord, est exigeante pour des personnes avec la schizophrénie. Elle est exigeante puisque l'environnement social ou professionnel susceptible de les embaucher n'est pas sensibilisé à la problématique de la maladie mentale, il ne souhaite pas en entendre parler ou fait comme si tout était normal. Steven, la même personne citée plus haut, ajoute :

> *Je croyais que tout le monde se souciait un peu de ça. Je me suis réveillé quand j'ai appris ça au travail* [que personne ne s'en soucie]. *Il y a quelques personnes là-bas, tu sais, qui ne s'en soucient pas du tout. Et quand tu penses à tout le chemin parcouru pour t'en sortir, ça refroidit. Ça, c'est un fait* […] *Ils ne m'apprécient pas, tu sais. Ils ne font rien pour mieux me connaître, ou ils ne veulent pas me connaître. Je ne sais vraiment pas pourquoi. Je ne sais pas.* (Jenkins et Carpenter-Song, 2008, p. 396)

Un contexte de travail comme celui qui est décrit par les personnes interrogées témoigne de difficultés ressenties par ces dernières et qui peuvent être vécues comme une source de détresse émotionnelle supplémentaire. Certaines personnes aux prises avec cette maladie ne sont peut-être pas prêtes à assumer cette situation ou à la vivre à nouveau. Elles peuvent choisir le retrait de toute activité professionnelle pour se protéger, mais ce retrait les met davantage en situation de dépendance, ce qui contrevient en quelque sorte aux valeurs dominantes de choix personnel, d'autonomie, d'indépendance et d'individualisme. Jenkins et Carpenter-Song (2008) suggèrent que le respect de ces valeurs est compromis par la symptomatologie psychotique.

Avec des données sociodémographiques qui comprennent des indicateurs culturels (Euro-Américains et Afro-Américains), ces auteurs mettent aussi en lumière l'influence du milieu ethnoculturel sur la perception de la maladie et des mécanismes sociaux différents de soutien et d'exclusion. Par exemple, les personnes ayant une schizophrénie provenant de la classe moyenne ou de franges plus scolarisées de la société semblent plus propices à souscrire à ces valeurs (de choix personnel, d'autonomie, d'indépendance et d'individualisme) que les classes subalternes, plus portées sur le soutien de la famille et de la communauté en général, y compris de l'Église locale.

CONCLUSION

En plus d'accéder à l'univers du sens et des représentations, la méthode ethnographique présente une option intéressante pour le chercheur qui souhaite comprendre la dynamique des interactions interpersonnelles dans un environnement donné. L'ethnographie n'exclut pas le recours à des mesures quantitatives (Bernard, 1995). Plusieurs méthodes d'analyse peuvent être employées en ethnographie : analyse de contenu, analyse narrative, analyse de discours, analyse par théorisation ancrée, analyse inspirée de la phénoménologie interprétative, etc. Les techniques d'enquête sont multiples : dialogues informels, entrevues semi-dirigées, entrevues ouvertes, récits de vie, groupes de discussion (*focus group*), observation directe, recours à l'audiovisuel, etc.

Même quand il s'agit d'étudier des phénomènes externes, celui qui utilise l'ethnographie comme méthode de recherche doit savoir que décrire un phénomène repose largement sur le positionnement théorique du chercheur, comme l'ont soutenu Emerson *et al*. (2011), pour qui la description est aussi un acte de perception et d'interprétation. En santé mentale et en réadaptation, l'observation de situations d'interactions *in situ* constitue sans doute le plus grand défi en matière de faisabilité, car les organisations ne vont pas nécessairement ouvrir leurs portes au chercheur ; c'est pourquoi il lui faut prendre le temps de tisser des liens privilégiés afin de faire sa place et de se faire accepter dans le milieu à l'étude. L'ethnographie n'est pas possible sans cette acceptation par le milieu. L'ethnographie demeure une approche très intéressante pour départager les éléments discursifs de l'expérience des éléments non linguistiques (gestes, expressions faciales, tenues corporelles, etc.) et pour leur donner un relief qui n'apparaît pas dans la seule entrevue. Elle permet aussi d'explorer, de décrire et d'analyser des situations qui échappent au regard quotidien. Elle peut contribuer ainsi à mieux cibler les interventions, celles qui sont les plus propices à être acceptées par les bénéficiaires ou, en revanche, celles qui sont susceptibles de rencontrer une résistance.

Pour ceux qui désirent pousser leur réflexion sur les fondements épistémologiques de l'ethnographie, voici quelques ouvrages à consulter. Pour des considérations plus générales qui vont de la conception du cadre théorique à la collecte des données, voir Spradley (1980) ; sur l'écriture ethnographique et la prise de notes de terrain, voir Emerson *et al*. (2011) et Sanjek (1996a) ; sur les paradigmes de la recherche qui peuvent guider un projet de recherche ethnographique, voir Silverman (2011) ; sur une manière de consigner et d'analyser les données ethnographiques, voir Hammersley et Atkinson (2007). En langue française, le chapitre 10 du recueil de Robert Mayer et Francine Ouellet (1991), intitulé « L'observation participante »,

offre une vue assez complète de cette approche : historique, technique d'observation, attitude du chercheur, insertion dans la communauté, unité d'observation, enregistrement des observations, etc.

RÉFÉRENCES

AGAR, M. (1982). « Toward an ethnographic language », *American Anthropologist*, vol. 84, p. 779-795.

AMERICAN PSYCHIATRIC ASSOCIATION – APA (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4e éd., Washington, American Psychiatric Press.

ANNELLS, M. (1996). « Grounded theory method : Philosophical perspectives, paradigm of inquiry, and postmodernism », *Qualitative Health Research*, vol. 6, p. 379-393.

BABISS, F. (2002). « Doing research in the field : Ethnographic methods », *Occupational Therapy in Mental Health*, vol. 18, p. 15-27.

BATESON, G. (1972). « Form and pathology in relationship », dans *In Steps to an Ecology of Mind : Collected Essays in Anthropology, Psychiatry, Evolution, and Epistemology*, San Francisco, Chandler Publication Company, p. 271-278..

BERNARD, R.H. (1995). *Research Methods in Anthropology : Qualitative and Quantitative Approaches*, Walnut Creek, AltaMira Press.

BLUMER, H. (1969). *Symbolic Interactionism : Perspective and Method*, Upper Saddle River, New Prentice-Hall.

BURR, V. (2003). *An Introduction to Social Constructionism*, Londres, Routledge.

CHWALISZ, K., S.R. SHAH et K.M. HAND (2008). « Facilitating rigorous qualitative research in rehabilitation psychology », *Rehabilitation Psychology*, vol. 53, p. 387-399.

CLIFFORD, J. (1986). « Introduction : Partial truths », dans J. Clifford et G. Marcus (dir.), *Writing Culture*, Berkeley, University of California Press, p. 1-26.

CORIN, E.E. (1990). « Facts and meaning in psychiatry : An anthropological approach to the lifeworld of schizophrenics », *Culture, Medicine and Psychiatry*, vol. 14, p. 153-188.

CORIN, E.E. et G. LAUZON (1994). « From symptoms to phenomenon : The articulation of experience in schizophrenia », *Journal of Phenomenological Psychology*, vol. 25, p. 3-50.

CÔTÉ, D. (2013). « Intercultural communication in health care. Challenges and solutions in work rehabilitation practices and training : A comprehensive review », *Disability and Rehabilitation*, vol. 35, p. 153-163.

COURTOIS, J.-P. et H. DESMET (2007). *Épistémologie et instrumentation en sciences humaines*, Wavre, Mardaga.

COUTU, M.-F., R. BARIL, M.-J. DURAND et A. ROULEAU (2007). « Representations : An important key to understanding workers' coping behaviors during rehabilitation and the return-to-work process », *Journal of Occupational Rehabilitation*, vol. 17, p. 522-544.

CRESWELL, J.W. (2007). « Five qualitative approaches to inquiry », dans J.W. Creswell, *Qualitative Inquiry and Research Design. Choosing Among Five Approaches*, Thousand Oaks, Sage Publications, p. 53-84.

CUCHE, D. (2001). *La notion de culture dans les sciences sociales*, Paris, La Découverte.

DELUZ, A. (2000). « Ethnopsychiatrie », dans A. Deluz (dir.), *Dictionnaire de l'ethnologie et de l'anthropologie*, Paris, Quadrige et Presses universitaires de France, p. 251-252.

EMERSON, R.M., R.I. FRETZ et L.L. SHAW (2011). *Writing Ethnographic Fieldnotes*, 2e éd., Chicago, University of Chicago Press.

GEERTZ, C. (1995). *After the Fact : Two Countries, Four Decades, One Anthropologist*, Cambridge, Harvard University Press.

GOFFMAN, E. (1961). *Asylums : Essays on the Social Situation of Mental Patients and Other Inmates*, New York, Anchor Books.

GOFFMAN, E. (1975). *Stigmate. Les usages sociaux des handicaps*, Paris, Éditions de Minuit.

GRATTON, D. (2009). *L'interculturel pour tous. Une initiation à la communication pour le troisième millénaire*, Anjou, Éditions Saint-Martin.

GRAVEL, S., H. DOUCET, D. LAUDY, L. BOUCHERON, A. BATTAGLINI, M.-È. BOUTHILLIER *et al.* (2009). « L'adhésion des populations immigrantes à la valeur de l'autonomie véhiculée en santé publique », *International Migration and Integration*, vol. 10, p. 31-47.

GROLEAU, D., A. YOUNG et L.J. KIRMAYER (2013). « The McGill Illness Narrative Interview (MINI) : An interview schedule to elicit meaning and modes of reasoning related to illness experience », *Transcultural Psychiatry*, vol. 43, p. 671-691.

HAMMERSLEY, M. et P. ATKINSON (2007). *Ethnography. Principles and Practice*, Londres et New York, Routledge.

IWAMA, M.K. (2003). « Toward culturally relevant epistemologies in occupational therapy », *The American Journal of Occupational Therapy*, vol. 57, p. 582-588.

JENKINS, J.H. et E.A. CARPENTER-SONG (2005). « The new paradigm of recovery from schizophrenia : Cultural conundrums of improvement without cure », *Culture, Medicine and Psychiatry*, vol. 29, p. 379-413.

JENKINS, J.H. et E.A. CARPENTER-SONG (2008). « Stigma despite recovery : Strategies for living in the aftermath of psychosis », *Medical Anthropology Quarterly*, vol. 22, p. 381-409.

KLEINMAN, A. (1980). *Patients and Healers in the Context of Culture : An Exploration of the Borderland between Anthropology, Medicine, and Psy*chiatry, Berkeley, University of California Press.

LAPERRIÈRE, A. (1995). « L'observation directe », dans B. Gauthier (dir.), *Recherche sociale. De la problématique à la collecte des données*, Québec, Presses de l'Université du Québec, p. 251-272.

LENCLUD, G. (2000). « Méthode ethnographique », dans A. Deluz (dir.), *Dictionnaire de l'ethnologie et de l'anthropologie*, Paris, Quadrige et Presses universitaires de France, p. 470-475.

MACEACHEN, E., A. KOSNY, S. FERRIER et L. CHAMBERS (2010). « The "toxic dose" of system problems : Why some injured workers don't return to work as expected », *Journal of Occupational Rehabilitation*, vol. 20, p. 349-366.

MALTERUD, K. (2001). « Qualitative research : Standards, challenges, and guidelines », *Lancet*, vol. 358, p. 483-488.

MARCUS, G. et M. FISCHER (1986). *Anthropology as Cultural Critique : An Experimental Moment in the Human Sciences*, Chicago, University of Chicago Press.

MASSÉ, R. (1995). « Qu'est-ce l'anthropologie de la santé ? », dans *Culture et santé publique*, Montréal, Gaëtan Morin, p. 15-46.

MAYER, R. et F. OUELLET (1991). « L'observation participante », dans R. Mayer et F. Ouellet (dir.), *Méthodologie de recherche pour les intervenants sociaux*, Montréal, Gaëtan Morin, p. 402-436.

MCREYNOLDS, C. J., L. C. KOCH et P. D. JR. RUMRILL (2001). Qualitative research strategies in rehabilitation, *Work,* vol. 16*,* p. 57-65.

MOLL, S., J.M. EAKIN, R.L. FRANCHE et C. STRIKE (2012). « When health care workers experience mental ill health : Institutional practices of silence », *Qualitative Health Research*, <doi : 10.1177/1049732312466296>.

MUCCHIELLI, A. (1998). *Approche systémique et communicationnelle*, Paris, Armand Colin.

NADER, L. (2012). « Ethnography as theory », *Journal of Ethnographic Theory*, vol. 1, p. 211-219.

NARAYAN, K. (2008). « Ethnography and fiction : where is the border ? », *Anthropology and Humanism*, vol. 24, p. 134-147.

POOREMAMALI, P., D. PERSSON et M. EKLUND (2011). « Occupational therapists' experience of working with immigrant clients in mental health care », *Scandinavian Journal of Occupational Therapy*, vol. 18, p. 109-121.

POPPER, K. (1959). *The Logic of Scientific Discovery*, Londres, Hutchinson.

SANJEK, R. (1996a). « A vocabulary for fieldnotes », dans R. Sanjek (dir.), *Fieldnotes. The Makings of Anthropology*, Ithaca, Cornell University Press, p. 92-121.

SANJEK, R. (1996b). *Fieldnotes. The Makings of Anthropology*, Ithaca, Cornell University Press.

SANJEK, R. (1996c). « On ethnographic validity », dans R. Sanjek (dir.), *Fieldnotes. The Making of Anthropology*, Cornell, Cornell University Press, p. 385-418.

SILVERMAN, D. (2011). « Ethnographic observation », dans *Interpreting Qualitative Data. A Guide to the Principles of Qualitative Research*, Los Angeles, Sage Publications, p. 111.

Sloots, M., J.H.M. Dekker, M. Pont, E.A.C. Bartels, J.H.B. Geertzen et J. Dekker (2010). « Reasons of drop-out from rehabilitation in patients of Turkish and Moroccan origin with chronic low back pain in The Netherlands : A qualitative study », *Journal of Rehabilitation Medicine*, vol. 42, p. 566-573.

Smith, D.E. (2002). « Institutional ethnography », dans T. May (dir.), *Qualitative Research in Action : An International Guide to Issues in Practice*, Londres, Sage Publications, p. 150-161.

Smith, D.E. (2005). « Knowing the social. An alternative design », dans *In Institutional Ethnography. A Sociology for the People*, New York, Altamira, p. 27-45.

Spencer, J., L. Krefting et C. Mattingly (1993). « Incorporation of ethnographic methods in occupational therapy assessment », *The American Journal of Occupational Therapy : Official Publication of the American Occupational Therapy Association*, vol. 47, p. 303-309.

Spradley, J.P. (1980). *Participant Observation*, Orlando, Harcourt Brace Jovanovich College Publishers.

Strathern, M. (1990). « Out of context : The persuasive fictions of anthropology », dans M. Manganero (dir.), *Modernist Anthropology*, Princeton, Princeton University Press, p. 80-130.

Symon, G. et C. Cassell (1998). « Reflections on the use of qualitative methods », dans *In Qualitative Methods and Analysis in Organizational Research. A Practical Guide*, Thousand Oaks, Sage Publications, p. 1-9.

White, B. (2012). « From experimental moment to legacy moment : Collaboration and the crisis of representation », *Collaborative Anthropologies*, vol. 5, p. 65-97.

Winkin, Y. (2001). *Anthropologie de la communication. De la théorie au terrain*, Paris, Seuil.

CHAPITRE 4

L'ÉTUDE DE CAS
Illustration d'une étude de cas multiples visant à mieux comprendre la participation au travail de personnes présentant un trouble de la personnalité limite

Kathy Dahl
Nadine Larivière
Marc Corbière

FORCES

- Elle permet de comprendre et d'analyser un phénomène d'intérêt complexe, en profondeur, en tenant compte de son contexte.
- Elle renvoie à une approche de recherche flexible recourant à la triangulation de sources de données multiples.
- Elle présente une forte validité interne.

LIMITES

- Elle peut générer une quantité importante de données et demander un temps de réalisation non négligeable.
- En raison de sa grande flexibilité, sa procédure de réalisation peut devenir complexe et confuse pour le chercheur novice.
- Sa validité externe peut poser problème et limiter la transférabilité des résultats.

L'étude de cas est une approche de recherche permettant l'étude d'un phénomène d'intérêt particulier (*le cas*) dans son contexte naturel et sans manipulation par le chercheur (Yin, 2009). Elle peut s'intéresser à un seul individu, à un groupe d'individus, à une communauté, à une institution ou à un événement (Hentz, 2012). À l'aide de méthodes de collecte de données multiples et de la triangulation de celles-ci, l'étude de cas génère une compréhension holistique, riche, détaillée et en profondeur du phénomène d'intérêt et de son contexte (Benbasat, Goldstein et Mead, 1987; Luck, Jackson et Usher, 2006).

L'étude de cas est utilisée dans plusieurs disciplines des sciences sociales et de la santé (Hentz, 2012). Elle peut avoir une finalité descriptive, exploratoire, explicative, ou encore être utilisée pour concevoir une théorie ou la tester (Gagnon, 2012; Hentz, 2012; Martinson et O'Brien, 2010). Elle est aussi souvent utilisée en évaluation de programmes (Martinson et O'Brien, 2010) (voir le chapitre 9 de cet ouvrage). Il importe de distinguer l'étude de cas comme approche de recherche de ses autres usages, tels que l'étude de cas pédagogique utilisée en enseignement ou encore l'étude de cas historique, qui permet de recueillir des données en lien avec un événement historique (Hentz, 2012). Ces types d'études de cas poursuivent d'autres objectifs que la recherche et ont des caractéristiques distinctes. Ainsi, seule l'étude de cas utilisée comme approche de recherche est abordée dans ce chapitre.

L'étude de cas est conceptualisée comme une approche plutôt que comme une méthode de recherche puisqu'en raison de sa flexibilité, elle n'exige pas une méthodologie unique de réalisation (Rosenberg et Yates, 2007). Le choix des techniques de collecte et d'analyse des données est tributaire de ce qui permettra de comprendre au mieux le phénomène en profondeur de même que de répondre, toujours de la meilleure façon possible, aux questions et objectifs de recherche (Rosenberg et Yates, 2007). Ceci en fait une approche intéressante pour sa flexibilité ainsi que pour la pluralité d'objets et de questions de recherche qu'elle permet d'étudier. Néanmoins, elle peut devenir complexe et difficile à conceptualiser pour le chercheur novice (Rosenberg et Yates, 2007).

Dans le cadre de ce chapitre, il sera d'abord question de décrire les différents types d'étude de cas. Ensuite, les étapes de réalisation de l'étude de cas seront présentées selon Gagnon (2012). Ainsi, la deuxième partie de ce chapitre illustrera la procédure de réalisation de l'étude de cas, et ce, dans le contexte d'un projet de recherche visant à comprendre la participation au travail des personnes présentant un trouble de la personnalité limite. Les différents choix méthodologiques de ce projet de recherche seront explicités, de la planification à la diffusion des résultats, afin de proposer une application pratique de l'étude de cas multiples.

1. CLASSIFICATION DES TYPES D'ÉTUDES DE CAS

Différentes conceptualisations (Stake, 1995 ; Yin, 2009) sont proposées pour la classification des études de cas, comme cela est présenté dans le tableau 4.1. Selon Stake (1995), les études de cas peuvent être classées selon trois catégories : *étude de cas intrinsèque*, *étude de cas instrumentale* et *étude de cas collectif*. Yin (2009), quant à lui, classifie les types d'étude de cas en deux grandes catégories : les *études de cas unique* et les *études de cas multiples*. Il apporte des précisions aux études de cas uniques, qu'il catégorise en fonction du but de l'étude. Yin ajoute également certains aspects méthodologiques à l'étude de cas multiples, soit la « réplication » du processus de recherche d'un cas à un autre, c'est-à-dire que chaque cas est étudié selon la même procédure et dans des conditions les plus similaires possibles.

Tableau 4.1.
Classification des types d'études de cas selon Stake (1995) et Yin (2009)

	Stake (1995)	Yin (2009)
Cas unique	▪ **Cas instrumental :** Compréhension en profondeur du cas dans un but de généralisation à un phénomène plus large. ▪ **Cas intrinsèque :** Compréhension en profondeur du cas pour ses caractéristiques particulières et sa singularité.	▪ **Cas critique :** Vérifier une théorie. ▪ **Cas extrême ou unique :** Décrire ou analyser un phénomène qui est rare. ▪ **Cas typique ou représentatif :** Étudier un cas considéré comme commun ou typique du phénomène à l'étude. ▪ **Cas révélateur :** Décrire un phénomène émergent. ▪ **Cas longitudinal :** Décrire l'évolution d'un phénomène à différents moments dans le temps.
Cas multiples	▪ **Étude de cas collectif :** Collecte de données auprès de plusieurs cas qui, par leur synthèse, permettront de comprendre le phénomène d'intérêt.	▪ **Étude de cas multiples :** Étude de plusieurs cas pour lesquels il y a « réplication » de la procédure de recherche.

2. ÉTAPES DE RÉALISATION DE L'ÉTUDE DE CAS

Il n'y a pas de standardisation du processus de réalisation de l'étude de cas (Hentz, 2012). Les écrits sur le sujet présentent souvent des descriptions sommaires ou incomplètes (Gagnon, 2012). Ainsi, les huit étapes proposées dans ce chapitre s'appuient sur les écrits de Gagnon (2012), qui suggère un processus de recherche détaillé, ancré dans l'application pratique de l'étude de cas et soucieux de la rigueur scientifique. Par ailleurs, le processus de recherche présenté par d'autres auteurs tels que Benbasat *et al.* (1987),

Martinson et O'Brien (2010), Merriam (1988), Rosenberg et Yates (2007), Stake (1995) et Yin (2009) s'inscrit également dans les étapes proposées par Gagnon.

2.1. Établir la pertinence de l'utilisation de l'étude de cas

D'abord, il importe de s'assurer que l'objet d'étude est approprié à l'étude de cas. Cette approche de recherche est indiquée lorsque l'objet d'étude est complexe, s'inscrit dans un contexte naturel dont il peut être difficilement dissocié et qui ne peut être manipulé par le chercheur (Walshe *et al.*, 2004 ; Yin, 2009). Par exemple, il pourrait être question d'étudier l'application du plan d'action en santé mentale (*objet d'étude*) dans un centre hospitalier du Nord-du-Québec (*contexte*). Dans cet exemple, on peut difficilement faire abstraction du contexte unique de cette région éloignée, lequel influence nécessairement l'objet qui est à l'étude. De plus, dans ce même exemple, le chercheur pourrait être intéressé à comprendre l'objet d'étude en considérant plusieurs points de vue tels que ceux des utilisateurs de services, des intervenants et des gestionnaires. L'étude de cas permet alors une compréhension complète du phénomène d'intérêt résultant de la triangulation des données recueillies selon les points de vue des divers acteurs impliqués (Walshe *et al.*, 2004). Enfin, l'étude de cas peut être pertinente dans des circonstances variées en fonction de ses diverses finalités : l'étude de phénomènes pour lesquels il y a peu de connaissances ou dont certains aspects demeurent inexplorés (finalité exploratoire), la description détaillée et complète du cas dans son contexte (finalité descriptive) ou l'exploration des relations de causalité (finalité explicative) (Gagnon, 2012 ; Hentz, 2012 ; Martinson et O'Brien, 2010).

2.2. Assurer la véracité des résultats

Assurer la validité et la fiabilité de l'étude de cas se fait de façon continue du tout début de l'étude jusqu'à la diffusion des résultats. Plusieurs stratégies s'avèrent utiles pour accompagner ce processus. Elles sont synthétisées dans le tableau 4.2 et sont détaillées dans les étapes subséquentes. Quelques informations additionnelles concernant le respect des critères scientifiques sont présentées dans les paragraphes suivants.

Tableau 4.2.
Stratégies pour assurer la rigueur scientifique de l'étude de cas

Critères de rigueur scientifique	**Stratégies pour assurer la rigueur scientifique**	**Étape du processus de recherche**
Validité interne	Comparer les résultats aux données de l'étude et aux écrits scientifiques.	Analyse
	Explorer les données et explications contradictoires.	Analyse
	Rédiger des notes de terrain et un journal de bord du chercheur.	Collecte Analyse
	Recourir à des sources de données multiples et à la triangulation des données.	Collecte Analyse
Validité externe	Éviter la sélection de cas trop particuliers.	Recrutement
	Sélectionner des cas reproductibles dans le temps.	Recrutement
	Utiliser un échantillonnage raisonné pour augmenter la variation de certaines caractéristiques des cas.	Recrutement
	Recruter plusieurs cas et effectuer une analyse intercas.	Recrutement Analyse
	Décrire en détail le contexte et les participants à l'étude.	Diffusion
Validité de construit	Définir clairement les concepts étudiés et leur opérationnalisation.	Préparation
	Utiliser des concepts et théories validés dans la littérature.	Préparation
	Sélectionner des cas permettant d'étudier le phénomène d'intérêt.	Recrutement
	Recourir à des sources de données multiples et à la triangulation des données.	Collecte Analyse
	Faire réviser le rapport par des pairs et les répondants.	Analyse Diffusion
	Décrire précisément la méthodologie utilisée.	Diffusion
Fiabilité	Documenter et décrire en détail chacune des étapes de la recherche.	Diffusion
	Impliquer des chercheurs externes à toutes les étapes de la recherche.	Toutes les étapes
	Enregistrer les données et en faciliter l'accès.	Collecte

Sources : Inspiré de Gagnon, 2012, et de Yin, 2009.

2.2.1. Stratégies pour assurer la validité

Selon Gagnon (2012, p. 30), « la validité interne est sans doute la principale force de l'étude de cas » en raison, d'une part, de stratégies de collecte de données s'appuyant sur la triangulation des informations recueillies et, d'autre part, d'un processus d'analyse visant la vérification continue des conclusions de l'étude. En ce qui concerne la validité externe, elle peut représenter la plus importante faiblesse de l'étude de cas et nécessite une attention particulière (Yin, 2009). En effet, comme elle s'intéresse à l'étude particulière d'un ou de quelques cas, la transférabilité des résultats en découlant peut être limitée. Les éléments à considérer à ce sujet concernent, notamment, la sélection des cas, qui sera discutée à l'étape 4 (recrutement des cas). Enfin, la validité de construit est assurée non seulement lors de l'étape 3 (préparation) par une définition claire et précise du cas étudié, mais aussi lors de l'étape 4 (recrutement des cas) par une sélection adéquate des cas (Gagnon, 2012).

2.2.2. Stratégies pour assurer la fiabilité

Ces stratégies visent à limiter l'influence de la subjectivité du chercheur sur les résultats de l'étude. Lors de l'étape 8 (diffusion des résultats), le chercheur doit décrire précisément sa démarche de recherche (Ghesquière, Maes et Vandenberghe, 2004). De plus, l'implication de plus d'un chercheur à toutes les étapes du processus de recherche favorise à nouveau la fiabilité des conclusions (Gagnon, 2012).

2.3. Préparation

Cette étape fait référence à ce qui doit être réalisé préalablement à toute action sur le terrain (Gagnon, 2012). La préparation implique trois sous-étapes : 1) formuler la question et les objectifs de recherche ; 2) définir le cas ; et 3) choisir le type approprié.

2.3.1. Formuler la question et les objectifs de recherche

Le chercheur consulte d'abord la littérature scientifique afin de cerner plus précisément le phénomène d'intérêt, de poser la question et de définir les objectifs de recherche (Gagnon, 2012). Peu importe sa finalité, l'étude de cas répond à des questions telles que « Comment ? » ou « Pourquoi ? » (Yin, 2009).

2.3.2. Définir le cas

Le chercheur définit, ensuite, de façon claire, précise et opérationnelle le « cas » étudié de même que son contexte et ses limites (Yin, 1999). Il s'agit de l'une des étapes les plus importantes de la réalisation de l'étude de cas, car elle affecte toutes les décisions méthodologiques ultérieures (Luck *et al.*, 2006 ; Rosenberg et Yates, 2007). Une définition imprécise peut conduire à la collecte de données inutiles et non pertinentes qui alourdiront le processus d'analyse (Luck *et al.*, 2006 ; Yin, 1999). De même, lors d'une étude de cas multiples, il pourrait en résulter une impossibilité de comparer les cas (Yin, 1999).

Définition du cas

Le cas où, tel que Yin (2009) le nomme, « *l'unité d'analyse* », peut être un individu, un groupe, une organisation, une communauté, un événement, un processus, une relation ou un projet ayant lieu dans une situation naturelle et réelle (Merriam, 1988 ; Yin, 2009). Le choix du cas est directement lié à la question de recherche (Merriam, 1988). Par exemple, si le chercheur désire étudier le processus de retour aux études d'une personne présentant un trouble psychotique, le cas ou l'unité d'analyse sera un individu présentant un trouble psychotique se retrouvant dans cette situation de vie. La littérature scientifique peut aider le chercheur à définir le cas ; il pourra s'appuyer sur des théories existantes ou des recherches antérieures ayant étudié le même phénomène et, plus tard, comparer ses résultats avec ceux d'autres études publiées (Rosenberg et Yates, 2007 ; Yin, 2009).

Définition des limites du cas

Les limites du cas sont déterminées par le chercheur, à nouveau en fonction de ses questions et objectifs de recherche. Les limites particulières du cas sont définies en fait de temps et d'espace, ainsi qu'en fonction de la nature même de l'objet d'étude (Sorin-Peters, 2004 ; Yin, 2009). Par exemple, s'il désire étudier une intervention particulière, le chercheur se posera des questions comme : « De quelle intervention s'agit-il ? », « À quel moment l'étude de cette intervention aura-t-elle lieu ? », « Quels intervenants y sont impliqués ? ». Les autres éléments comme l'organisation, le système de santé et les acteurs ne faisant pas partie de cette intervention, mais y exerçant une influence sont alors considérés comme faisant partie du contexte de l'étude de cas.

2.3.3. Choisir le type approprié

La nature du phénomène à l'étude, la question et les objectifs de recherche, ainsi que les ressources et le temps disponibles guideront le chercheur dans son choix de l'étude de cas unique ou de l'étude de cas multiples. Ces informations sont décrites dans le tableau 4.3.

Tableau 4.3.
Choisir entre l'étude de cas unique et de cas multiples

	Étude de cas unique	Étude de cas multiples
Buts	▪ Tester une théorie déjà reconnue (Gagnon, 2012). ▪ Étudier un phénomène unique, extrême ou encore peu exploré à ce jour (Yin, 2009). ▪ Étudier un phénomène de façon longitudinale (Yin, 2009).	▪ Décrire un phénomène, générer ou vérifier une théorie (Benbasat *et al.*, 1987). ▪ Produire une compréhension plus générale d'un phénomène (Benbasat *et al.*, 1987). ▪ Souligner les similitudes ou les différences entre plusieurs cas (Gagnon, 2012). ▪ Étudier des cas se produisant généralement dans des situations variables (Gagnon, 2012).
Avantages	▪ Volume de données plus restreint, donc la collecte et l'analyse de données demandent moins de temps et de ressources (Yin, 2009).	▪ Résultats considérés plus rigoureux et représentatifs (Yin, 2009). ▪ Assurance d'une meilleure confidentialité et de l'anonymat des sujets participant à l'étude (Gagnon, 2012).
Inconvénients	▪ Transférabilité des résultats limitée (Yin, 2009).	▪ Compréhension plus superficielle de chacun des cas (Yin, 2009). ▪ Exigence de ressources et de temps importants due au volume considérable de données recueillies (Gagnon, 2012). ▪ Le phénomène à l'étude doit permettre la réplication du processus de recherche à chacun des cas (Yin, 2009).

2.3.4. Élaborer le protocole de recherche

Le chercheur rédige le protocole de recherche en prévoyant la manière dont seront sélectionnés et recrutés les cas, ainsi que les méthodes de collecte et d'analyse des données. Bien que le protocole soit rédigé avant toute démarche sur le terrain, il ne faut pas oublier que l'étude de cas utilisée en recherche est une approche flexible et peut être adaptée au fur et à mesure afin de tenir compte des nouvelles données (Yin, 2009).

2.4. Recrutement des cas

L'échantillonnage raisonné, qui repose sur une sélection de participants selon des critères d'inclusion et d'exclusion précis, représente la méthode de choix pour l'étude de cas. En effet, l'échantillon raisonné permet le recrutement de participants pertinents au phénomène à l'étude, à partir desquels il sera possible d'en avoir une compréhension plus approfondie (Creswell *et al.*, 2007; Fortin, 2010; Stake, 1995). Par exemple, si l'objectif de l'étude est de décrire l'utilisation d'une intervention spécialisée, il sera nécessaire de recruter une clinique où cette intervention est implantée. Pour l'étude de cas multiples, l'échantillonnage raisonné permet également la sélection de cas homogènes ou variables en fonction de l'objectif de recherche (Gagnon, 2012).

En ce qui concerne plus particulièrement l'étude de cas multiples, d'autres critères doivent également être considérés. D'abord, les cas recrutés doivent permettre la *réplication* de la procédure de recherche d'un cas à l'autre (Yin, 2009). Aussi, ils doivent permettre la *comparaison* à d'autres situations et, donc, requérir des caractéristiques particulières, mais toutefois semblables à d'autres cas de la réalité (Gagnon, 2012). Ensuite, il convient de choisir des cas *reproductibles* dans le temps afin de pouvoir, une nouvelle fois, transférer les résultats à d'autres situations (Gagnon, 2012). De plus, il importe de déterminer le nombre nécessaire de cas à recruter, ce qui peut être différent d'une étude à une autre (Creswell *et al.*, 2007). Le *nombre de cas* suffisant est déterminé en fonction du nombre de réplications qui permettra d'accorder une plus grande confiance aux résultats obtenus: plus il y a de variables importantes pouvant influencer les résultats, plus il apparaît nécessaire d'étudier un nombre élevé de cas (Yin, 2009). L'étude d'au moins trois cas offre une variation intéressante, mais, de façon générale, le chercheur tentera de recruter de six à neuf cas (Martinson et O'Brien, 2010). Néanmoins, il ne faut pas perdre de vue que l'étude de plusieurs cas générera une quantité considérable de données, ce qui sera plus coûteux en temps et en argent et pourra limiter la profondeur de la compréhension qui en résultera (Gagnon, 2012). Par conséquent, il est déconseillé d'étudier plus de quinze cas (Martinson et O'Brien, 2010). Enfin, en cours d'étude, le chercheur peut décider de cesser le recrutement de nouveaux cas avant le nombre initialement prévu lors de l'atteinte de la saturation des données, c'est-à-dire lorsque les nouveaux cas n'apporteront plus de nouvelles informations (Fortin, 2010).

2.5. Collecte des données

L'utilisation de sources de données multiples est l'une des caractéristiques principales de l'étude de cas (Yin, 2009). En appuyant les résultats sur plus d'une source de données et par leur triangulation, l'étude de cas permet d'assurer une compréhension en profondeur et valide du phénomène à l'étude ainsi que de son contexte (Rosenberg et Yates, 2007 ; Yin, 2009). L'étude de cas offre donc l'avantage d'aller au-delà de la simple analyse individuelle d'une ou de plusieurs sources de données. Néanmoins, cet aspect fait de la collecte de données une étape exigeante en temps et nécessite une connaissance adéquate des diverses techniques de collecte de données (Yin, 2009). Les différentes sources de données utilisées pour l'étude de cas sont décrites dans le tableau 4.4.

Tableau 4.4.
Description des différentes sources de données utilisées dans l'étude de cas

Source de données	Description	Objectifs
L'entrevue	L'entrevue peut être libre, semi-dirigée ou directive. C'est souvent la source de données la plus importante dans une étude de cas.	Permet de recueillir le point de vue et la perception des répondants.
L'observation participative ou directe	L'observation peut être méthodique à l'aide d'un protocole d'observation ou informelle en recueillant les informations lors de la visite d'un lieu ou pendant les entrevues.	Vise à recueillir des informations sur des éléments observables (comportements et environnement).
L'observation participante	Le chercheur joue un rôle actif dans la situation qui est à l'étude. Par exemple, le chercheur est membre de la communauté ou de l'organisation qui est à l'étude.	Permet d'avoir accès à des informations qui ne seraient pas disponibles par une observation directe.
L'analyse documentaire	Analyse de documents administratifs, de lettres, de publications ou de journaux personnels relatifs aux participants à l'étude.	Permet de confirmer les informations obtenues par d'autres sources et de recueillir de nouvelles informations sur le contexte.
L'examen de documents d'archives	Informations relatives au recensement, à des sondages, aux statistiques d'une organisation ou d'un service (p. ex., nombre de clients ou de professionnels, budget).	
L'évaluation d'artéfacts physiques	Des outils, des instruments, des œuvres d'art ou tout élément physique relatif à l'objet d'étude.	

Source : Adapté de Yin, 2009.

Aussi, le chercheur note régulièrement ses observations, ses impressions, ses réflexions sur la procédure méthodologique et ses hypothèses lors de la collecte des données (p. ex., notes de terrain). Ces notes permettent d'ajuster la collecte de données tout au long de l'étude et fournissent des informations complémentaires qui sont utiles lors de l'analyse des données (Gagnon, 2012).

2.6. Analyse des données (traitement)

L'analyse des données est sans aucun doute l'étape la plus complexe de l'étude de cas (Yin, 2009). Elle débute dès le commencement de la collecte des données. Ces deux étapes (collecte et analyse) se déroulent de manière itérative tout en s'influençant mutuellement (Hanley-Maxwell, Hano et Skivington, 2007; Sorin-Peters, 2004).

Devant le volume important d'informations, il convient, d'abord, d'épurer et d'organiser les données qui seront pertinentes pour l'étude (Miles et Huberman, 2003). Comme l'étude de cas comporte différentes sources de données, celles-ci sont d'abord transcrites dans leur intégralité. Ensuite, le chercheur effectue la codification thématique des données recueillies (attribution d'un code à des segments de texte constituant des unités de signification). Cette codification servira à extraire les informations, relations et significations des données sous une forme plus facile à analyser (Miles et Huberman, 2003). La conceptualisation, le développement d'hypothèses ainsi que la triangulation des données peuvent être réalisés par l'extraction des données sous la forme de tableaux, de schémas ou de graphiques (Miles et Huberman, 2003). Enfin, cette étape peut être soutenue par l'utilisation de logiciels d'analyse (p. ex., NVivo, QDA Miner) (Gagnon, 2012).

2.7. Analyse des données (interprétation)

Le chercheur ou l'équipe de recherche tente ensuite de générer une explication théorique plus générale du phénomène, puis à la vérifier à l'aide de la triangulation des données afin d'assurer sa validité (Gagnon, 2012). L'analyse est alors poussée à un niveau plus abstrait pour dégager les thèmes généraux et les relations possibles entre les différents concepts du phénomène à l'étude (Sorin-Peters, 2004). Pour ce faire, une analyse individuelle de chaque cas (intracas ou «*within case*») est conduite et, dans le cas d'une étude de cas multiples, une analyse entre les cas est par la suite entreprise (intercas ou «*cross-case*») (Ghesquière *et al.*, 2004; Yin, 2009). L'*analyse intracas* vise à dégager les tendances et à recontextualiser le phénomène afin de mener à une compréhension en profondeur pour chacun des cas

(Gagnon, 2012; Rosenberg et Yates, 2007). Dans l'*analyse intercas*, les cas sont comparés entre eux pour en faire ressortir les divergences et les similarités. Le chercheur tente alors de déceler et d'expliquer ces différences et ces similitudes, puis explore les liens de *causalité* possibles (Ghesquière *et al*., 2004; Sorin-Peters, 2004).

En vue de produire des conclusions valides, le chercheur utilise différentes stratégies d'analyse. D'une part, le recours à la triangulation des sources de données permet d'assurer que les conclusions sont soutenues par les différentes sources (Martinson et O'Brien, 2010). D'autre part, le chercheur peut comparer ses conclusions aux données de l'étude (entre les cas), ainsi qu'aux résultats d'études publiées dans la littérature scientifique (Yin, 2009). Dans la mesure où des informations présentent des contradictions, le chercheur est tenu de vérifier si elles ne sont pas le produit d'erreurs méthodologiques et, le cas échéant, de tenter de les expliquer et d'en tenir compte dans l'analyse finale (Gagnon, 2012).

2.8. Diffusion des résultats

La diffusion des résultats d'une étude de cas prend la forme d'un rapport (« *case study report* »). Ce dernier doit mettre en lumière ce qui a été appris du ou des cas étudiés (Creswell *et al*., 2007). De façon générale, ce rapport, comme tout rapport de recherche, décrit de façon transparente et détaillée les étapes qui ont été suivies (Martinson et O'Brien, 2010). De plus, le contexte de chaque cas est décrit, ce qui permet au lecteur d'évaluer l'applicabilité des résultats à d'autres situations (Ghesquière *et al*., 2004). Afin d'illustrer et d'appuyer les conclusions, des extraits de témoignages (*verbatim*) des répondants doivent aussi être présents dans le rapport (Gagnon, 2012). Il est également recommandé de vérifier le contenu du rapport auprès d'autres chercheurs de l'équipe, mais également auprès de participants à l'étude de cas afin d'en assurer la validité (Martinson et O'Brien, 2010; Yin, 2009).

Le rapport peut prendre différentes formes qui sont tributaires à la fois du type de l'étude de cas et des lecteurs ciblés (Yin, 2009). Pour une *étude de cas unique*, le chercheur rédige le récit détaillé du cas et présente les conclusions selon les différents thèmes émergeant de l'analyse (Martinson et O'Brien, 2010). Dans le cas d'une *étude de cas multiples*, différents choix pour la rédaction du rapport sont possibles (Yin, 2009) : 1) la description individuelle de chacun des cas, puis une synthèse de tous les cas selon les différents thèmes émergeant de l'analyse intercas; et 2) la synthèse des thèmes émergeant de l'analyse intercas seulement sans aucune description individuelle des cas. Par ailleurs, lorsque la synthèse de tous les cas est l'objectif premier de l'étude, il n'est pas nécessaire de rédiger le rapport sous un format de cas individuels (Yin, 1981).

3. ILLUSTRATION DE L'ÉTUDE DE CAS

Cette partie est une illustration pratique de l'étude de cas multiples, réalisée dans le cadre d'un projet de recherche qui porte sur la participation au travail de personnes présentant un trouble de la personnalité limite (TPL).

Le trouble de la personnalité limite est défini comme un mode général d'instabilité des relations interpersonnelles, de l'image de soi et des affects, avec une impulsivité marquée qui apparaît au début de l'âge adulte (APA, 2000). À ce jour, il est documenté que la participation au travail est la sphère du fonctionnement la plus touchée dans ce groupe (Cramer, Torgersenet et Kringlen, 2006 ; Larivière *et al.*, 2010 ; Skodol *et al.*, 2002). Néanmoins, peu d'écrits font état de façon approfondie des répercussions du TPL sur la participation au travail (réinsertion, retour et maintien au travail). L'objectif de cette étude est de combler cette lacune par le recours à l'étude de cas comme approche de recherche.

3.1. Établir la pertinence de l'utilisation de l'étude de cas

D'abord, une recension des écrits a permis de définir le phénomène à l'étude au moyen du cadre théorique de la participation au travail des personnes aux prises avec un trouble mental (Corbière et Durand, 2011). Selon ce cadre théorique, la participation au travail fait référence à quatre grandes situations socioprofessionnelles : *a*) réinsertion au travail (personne en invalidité qui est en rupture de travail et qui est inscrite dans un processus de réinsertion) ; *b*) retour au travail (personne en congé de maladie qui a conservé un lien d'emploi et qui est inscrite dans un processus de retour au travail) ; *c*) maintien au travail (personne qui a gardé un emploi après une réinsertion ou un retour au travail ou qui est restée active sur le marché du travail malgré son problème de santé) ; *d*) insertion au travail (personne en invalidité qui s'insère pour la première fois dans le marché du travail). Dans ce cadre théorique, les interactions du système personnel et sociétal (caractéristiques de la personne, de la famille et des amis), du système organisationnel (milieu de travail, collègues, supérieur immédiat, employeur, syndicat), du système de santé (soins de santé, équipe traitante) et du système de l'assurance (agent de réadaptation, conseiller en emploi, compagnie d'assurance) influencent la participation au travail de la personne ayant un trouble mental. Autrement dit, l'interaction des divers facteurs, issus de chacun de ces systèmes et s'inscrivant dans un contexte socioéconomique et culturel particulier, peut ainsi faciliter la participation au travail d'une personne présentant un trouble mental, notamment un TPL, ou y faire obstacle.

Ce modèle a ainsi permis de définir le phénomène d'intérêt, soit la participation au travail, et il convenait alors de déterminer s'il était approprié à l'étude de cas. D'abord, la participation au travail est un phénomène complexe : elle se définit selon quatre situations socioprofessionnelles, implique différents acteurs interagissant entre eux, et le tout ne peut être dissocié de son contexte naturel (les lois, les politiques et le contexte socio-économique). Ensuite, il apparaît important d'étudier le phénomène dans son contexte actuel, sans manipulation par le chercheur, afin de bien cerner tous les facteurs en cause. De plus, la recension des écrits a fait ressortir des lacunes dans la compréhension de la participation au travail chez les personnes présentant un TPL, ce qui souligne le caractère émergent de ce phénomène d'intérêt. Enfin, l'expérience clinique a mis en évidence l'importance de considérer aussi la perception des intervenants dans la compréhension du phénomène. Pour l'ensemble de ces raisons, il est donc apparu judicieux d'avoir recours à l'étude de cas pour mieux comprendre la participation au travail des personnes présentant un TPL.

3.2. Assurer la véracité des résultats

Plusieurs stratégies ont été mises en œuvre, tout au long de l'étude, afin d'assurer la validité et la fiabilité, telles que l'implication de trois chercheurs à toutes les étapes du processus de recherche. Ces stratégies sont décrites dans les sections subséquentes détaillant ce qui a été réalisé lors de chacune des étapes de ce projet de recherche.

3.3. Préparation

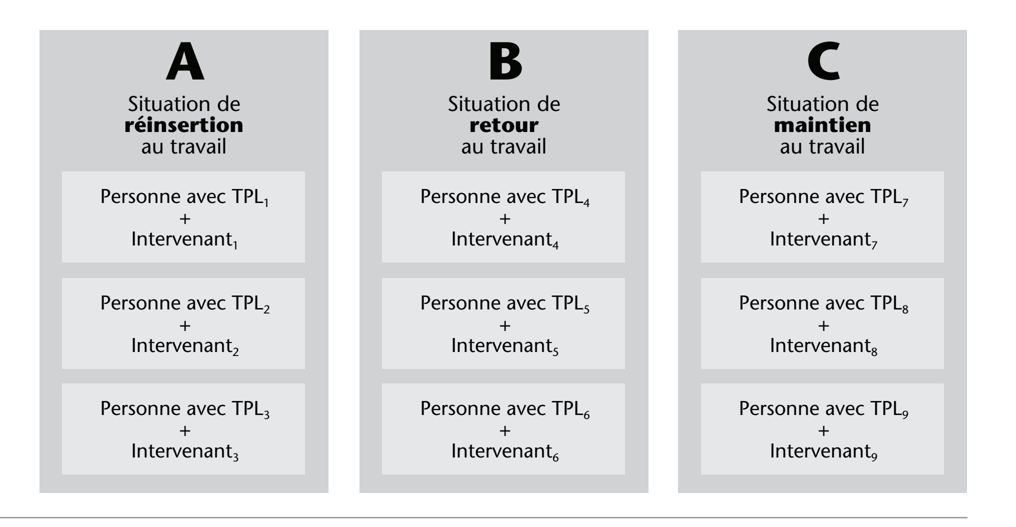
3.3.1. Formuler la question et les objectifs de recherche

La recension des écrits a permis de cibler deux questions de recherche : quels sont les obstacles et les facilitateurs de la participation au travail des personnes présentant un TPL et comment ces facteurs influencent-ils la participation au travail ? Ces questions correspondent à un questionnement de type « Comment ? », lequel est cohérent avec les finalités descriptives et explicatives de l'étude de cas (Yin, 2009). De ces deux questions a découlé l'objectif de cette étude : identifier, décrire et comprendre les obstacles et les facilitateurs de la participation au travail des personnes présentant un TPL en considérant trois situations socioprofessionnelles : *a*) réinsertion au travail, *b*) retour au travail et *c*) maintien au travail. L'insertion au travail n'a pas été étudiée puisque, dans le milieu clinique où avait lieu l'étude, cette situation socioprofessionnelle était peu fréquente, voire absente.

3.3.2. Définir le cas

Le cadre théorique de la participation au travail des personnes aux prises avec un trouble mental (Corbière et Durand, 2011) a donc été utilisé afin de définir le cas, ses limites et son contexte. Il a permis de cerner précisément le phénomène à l'étude et d'asseoir ce projet de recherche sur des bases solides. Chaque cas ou unité d'analyse a été conçu selon une dyade personne avec un TPL et son intervenant, et ce, dans l'une des trois situations socioprofessionnelles de la participation au travail (figure 4.1). Cette définition du cas a permis une compréhension en profondeur du phénomène à l'étude, résultant de l'analyse du point de vue de deux acteurs différents. Ainsi, pour chacune des situations socioprofessionnelles, trois cas ou dyades (personne présentant un TPL et son intervenant) ont été étudiés, pour un total de neuf dyades (ou dix-huit participants). Afin de délimiter clairement le cas, seule la situation socioprofessionnelle actuelle d'une personne, au moment de la collecte des données, a été considérée pour cette recherche.

Figure 4.1.
Illustration des cas (ou unités d'analyse) dans l'étude de la participation au travail de personnes présentant un TPL



3.3.3. Choisir le type approprié

Pour cette recherche, l'étude de cas multiples a été choisie, car elle permettait l'étude de trois situations socioprofessionnelles du cadre théorique de la participation au travail. De plus, l'étude de trois cas par situation

socioprofessionnelle a permis une compréhension approfondie et fine de chacune d'entre elles. Par ailleurs, le phénomène d'intérêt se prêtait aisément à la réplication dans le temps, avec l'usage d'une procédure similaire pour tous les cas.

3.3.4. Élaborer le protocole de recherche

Le protocole de recherche a ensuite été rédigé afin de décrire les méthodes que la chercheuse mettrait en place pour l'étude du phénomène. Tous les choix méthodologiques ont fait l'objet d'une réflexion approfondie de la part de l'équipe de recherche.

3.4. Recrutement des cas

Il a été décidé par l'équipe de recherche de sélectionner et de recruter les cas par échantillonnage raisonné afin d'optimiser l'analyse des résultats. Pour ce faire, des critères de sélection ont été établis, et ce, pour les deux catégories d'acteurs faisant partie de cette étude : les personnes présentant un TPL et leurs intervenants. Par exemple, l'un des critères d'exclusion, pour les personnes présentant un TPL, était la présence d'une incapacité au travail due à un trouble physique, sensoriel ou à une déficience intellectuelle. De plus, l'échantillonnage raisonné a permis d'augmenter la variabilité de certaines caractéristiques des participants à l'étude (p. ex., le sexe, l'âge, la profession, le type d'intervenant et les services de santé reçus). Tous ces éléments ont permis de soutenir la validité, la fiabilité et la transférabilité des résultats par le recrutement de cas pertinents et diversifiés (Gagnon, 2012).

En ce qui concerne le nombre de cas pour le recrutement, il a été fixé à neuf dyades (TPL-intervenant) pour un total de dix-huit participants. Par la réplication de la méthode de recherche à trois cas pour chacune des situations de participation au travail, il était alors possible de les comparer, d'en faire une synthèse et ainsi d'augmenter notre confiance quant à l'interprétation des résultats obtenus (Martinson et O'Brien, 2010). De plus, la variabilité des caractéristiques des participants étant souhaitée, il apparaissait nécessaire d'étudier plus d'un cas pour chaque situation de participation au travail (Yin, 2009).

3.5. Collecte des données

Pour cette étude, la principale méthode de collecte de données était l'*entrevue*. Deux entrevues semi-dirigées par participant ont été réalisées. Deux guides d'entrevue, comportant principalement des questions ouvertes, ont été élaborés à la suite d'une recension des écrits ; le premier pour les personnes présentant un TPL et le second pour les intervenants. Des exemples de questions sont présentés dans le tableau 4.5. Le guide d'entrevue a évolué au fur et à mesure de la collecte des données afin d'inclure les éléments émergeant en cours d'étude (Gagnon, 2012). La première entrevue a été effectuée en face à face et enregistrée sur bande audionumérique. La seconde a été réalisée par téléphone, deux à six mois après la première, en vue de clarifier, d'approfondir et de valider certains éléments de la première entrevue.

Tableau 4.5.
Exemples de questions du guide d'entrevue (première entrevue) dans l'étude de la participation au travail de personnes présentant un TPL

Entrevue avec les personnes présentant un TPL	▪ Pouvez-vous me parler de la manière dont se déroule votre réinsertion au travail/retour au travail/maintien au travail ? ▪ Pouvez-vous me parler des éléments qui vous aident dans votre réinsertion au travail/retour au travail/maintien au travail ? ▪ Pouvez-vous me parler des éléments faisant obstacle à votre réinsertion au travail/retour au travail/maintien au travail ?
Entrevue avec les intervenants	▪ Pouvez-vous me parler de la manière dont se déroule la réinsertion au travail/retour au travail/maintien au travail de cette personne ? ▪ Selon vous, quels sont les éléments qui ont aidé la réinsertion au travail/retour au travail/maintien au travail de cette personne ? ▪ Selon vous, quels sont les éléments faisant obstacle à la réinsertion au travail/retour au travail/maintien au travail de cette personne ?

Au moment de la première entrevue, un *questionnaire sociodémographique* était également rempli par les participants dans le but de brosser leur portrait, de documenter leur contexte et d'obtenir des informations additionnelles pour l'analyse des résultats (p. ex., le parcours professionnel de chaque participant). Des *notes de terrain* et un *journal de bord de la chercheuse* ont été rédigés pendant tout le déroulement de l'étude. Ces notes incluaient des observations, des réflexions sur la procédure méthodologique (p. ex., des modifications à apporter aux guides d'entrevue), des hypothèses, des critiques (p. ex., sur le déroulement des entrevues), des idées, des questionnements et le ressenti de la chercheuse, sans aucune censure. Ceci a permis à la chercheuse de tenir compte de sa subjectivité et de ses préconceptions et ainsi de favoriser un processus de recherche rigoureux (Gagnon, 2012).

Aussi, certains *documents* ont été consultés afin de comprendre le contexte de chaque situation socioprofessionnelle (p. ex., les lois et politiques gouvernementales québécoises relatives à l'incapacité au travail, la description des services de santé reçus par les participants) et de recontextualiser les résultats lors de l'analyse.

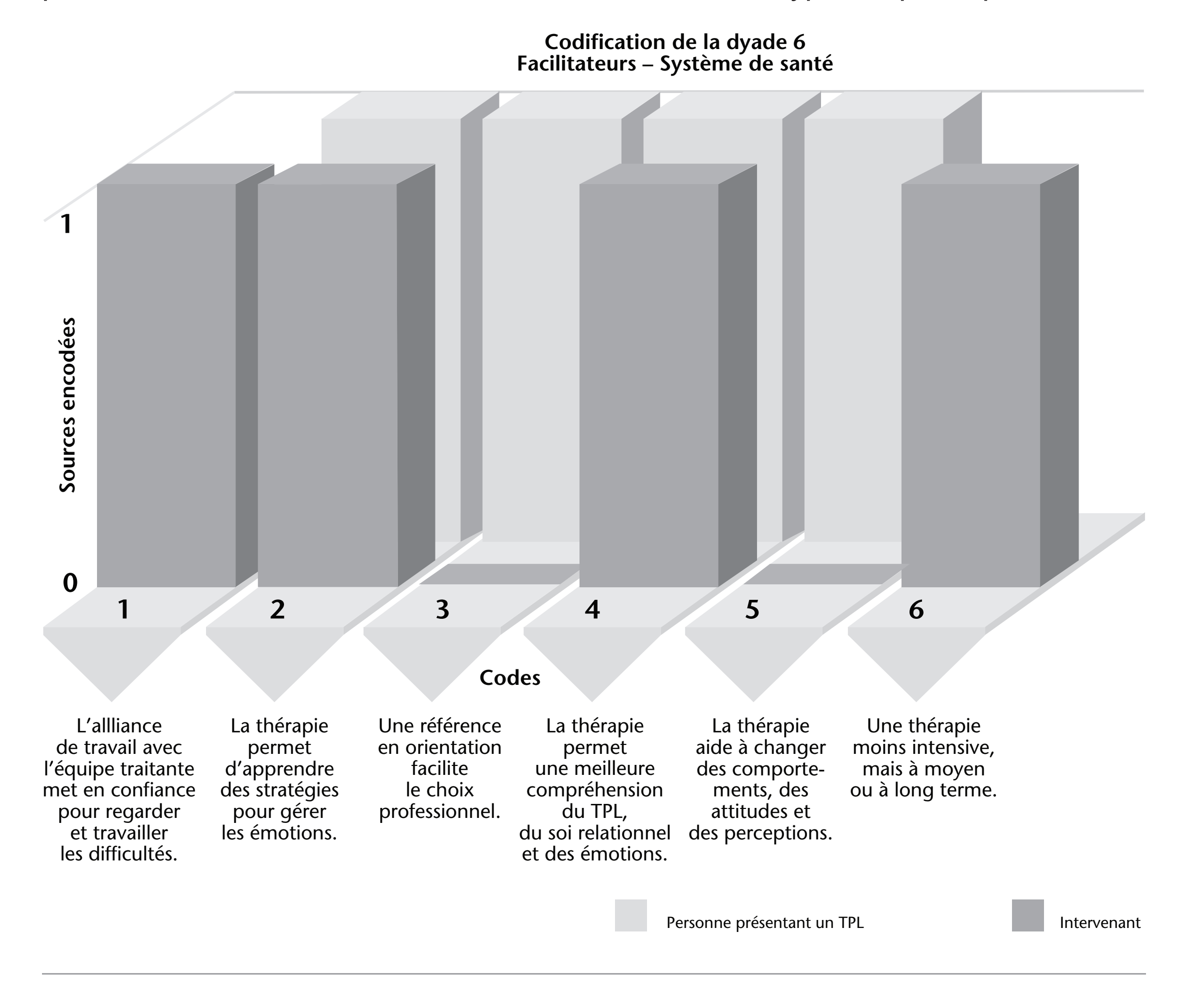
3.6. Analyse des données (traitement)

Le contenu des entrevues a été transcrit dans son intégralité et un *guide de codes* a été élaboré afin de systématiser l'analyse des données. Ce guide décrit les différentes catégories (codes) qui ont été utilisées ainsi que leur définition. Trois chercheurs ont été impliqués dans la création de ce guide afin d'assurer la validité de la codification et d'analyser les entrevues. Comme cette étude incluait l'analyse de plusieurs cas, le *logiciel* NVivo (version 9.0) a été utilisé. L'utilisation de ce logiciel a facilité l'organisation des données, la rédaction des mémos et la vérification du processus d'analyse par l'équipe de recherche. À nouveau, les trois chercheurs de l'équipe ont été impliqués dans cette étape afin d'effectuer le *contre-codage* des entrevues, en discuter et ajuster la codification. Par exemple, les entrevues de la première dyade ont été codées entièrement par les trois chercheurs. Puis, les entrevues des deux dyades suivantes ont été codées par deux chercheurs. Le reste de la codification, effectuée par un seul chercheur, a été vérifiée par l'un des deux autres chercheurs de l'équipe. Des discussions entre les membres de l'équipe de recherche ont eu lieu régulièrement afin d'ajuster la codification et d'obtenir une fidélité interjuges satisfaisante (80 %). Un *journal de mémos* a également été rédigé afin de noter les réflexions qui ont émergé en cours d'analyse.

3.7. Analyse des données (interprétation)

La production de *schémas explicatifs*, de *matrices* et de *graphiques* à l'aide du logiciel NVivo (version 9.0) a permis d'illustrer les différences et les similitudes au cours de l'analyse intracas et intercas. La figure 4.2 présente un graphique réalisé à l'aide du logiciel NVivo (version 9.0), issu de l'analyse intracas des facilitateurs du système de santé de la dyade 6. Cette représentation visuelle permet une comparaison rapide entre les éléments nommés par les deux types de participants à cette étude. Ainsi, on peut voir que les codes : « La thérapie permet d'apprendre des stratégies pour gérer les émotions » et « La thérapie permet une meilleure compréhension du TPL, du soi relationnel et des émotions » ont été tous deux mentionnés par l'intervenant et la personne avec un TPL de cette dyade.

Figure 4.2.
Exemple de graphique soutenant le processus d'analyse intracas pour observer les similitudes et les différences entre les types de participants



Pour chacun des cas pris individuellement, une analyse intracas a été effectuée afin d'identifier et d'expliquer les facteurs faisant obstacle et facilitant la participation au travail. Le contenu de cette analyse a été validé auprès de chacun des participants lors de la deuxième entrevue. Par la suite, une analyse intercas a été effectuée afin de comparer : 1) les trois dyades à l'intérieur de chacune des situations socioprofessionnelles ; 2) les trois situations socioprofessionnelles entre elles ; et 3) les différents points de vue (intervenants et personnes avec un TPL). Des thèmes généraux explicatifs

ont été élaborés en explorant les divergences et similitudes. Les propositions explicatives générées ont pu être confirmées ou infirmées par la triangulation des données et la comparaison des résultats avec la littérature scientifique (Yin, 2009). Les données sociodémographiques et les documents recueillis ont été utilisés afin de mettre en contexte les résultats obtenus.

3.8. Diffusion des résultats

Un article de recherche est en cours de rédaction pour rendre compte de cette étude. Une attention particulière a été portée à la description des méthodes, des caractéristiques des participants à l'étude ainsi que du contexte (Gagnon, 2012). Comme il s'agissait d'une étude de cas multiples visant principalement à faire une synthèse des cas, l'information a été organisée en fonction des thèmes généraux émergeant de chacune des situations de participation au travail. Les résultats ont été illustrés par des extraits de témoignages de divers participants. Le tableau 4.6 donne un exemple d'un thème issu de cette analyse.

Tableau 4.6.
Exemple de résultats dans l'étude de la participation au travail de personnes présentant un TPL

OBSTACLE : Hypersensibilité et préoccupation constante des perceptions de l'autre sur soi qui portent à la méfiance dans les relations de travail.

	Personne présentant un TPL	Intervenant
Réinsertion au travail	« *Je me méfie d'eux* [les collègues de travail]. *Je me demande ce qu'ils pensent de moi. Je me demande de quoi j'ai eu l'air. Je me demande s'ils ne m'aiment pas. Je me demande s'ils n'ont pas parlé dans mon dos.* [...] *J'ai toujours besoin d'analyser tout ce que je dis, ce qu'ils regardent, ce qu'ils pensent.* » (Personne avec un TPL, dyade 1)	Aucun témoignage ne touchait à ce code.
Retour au travail	« *La méfiance... J'ai beaucoup de méfiance ! Je le vois particulièrement dans ce travail-là.* [...] *On dirait que je me suis créé une méfiance et que je n'ai presque plus confiance en personne dans l'usine.* [...] *J'ai l'impression que je me sens jugé. Cette méfiance-là fait en sorte que je n'ai pas envie de travailler avec les gens à côté de moi. Je n'ai pas envie d'être en contact avec certaines personnes parce que j'ai l'impression qu'elles me détestent ou qu'elles me jugent ou qu'elles ne m'aiment pas ou etc. Donc je n'ai pas envie du tout d'être en contact avec certaines personnes*. » (Personne avec un TPL, dyade 6)	« *Il a eu certaines difficultés dans le sens où il pensait que les gens parlaient de lui. Je ne dirais pas comploter, mais parlaient dans son dos. Alors, il avait moins confiance en ses collègues de travail. Il avait l'impression qu'on le voyait d'un mauvais œil, qu'on le dénigrait, qu'on parlait dans son dos. Donc, ceci apparaissait l'influencer dans le sens où il était trop vigilant. Il était vraiment hypersensible. Puis, il a cédé. Il avait trop de pression par rapport à ça.* » (Intervenant, dyade 6)
Maintien au travail	« *Mais en même temps moi je suis méfiante d'elle* [une collègue de travail] *maintenant. Ça veut dire que j'évite d'être dehors en même temps qu'elle.* [...] *Puis, lundi, elle vient me parler : "Bonjour, ça va bien ?" Super gentille. Moi cette attitude-là, j'ai une méfiance incroyable. Tu me piques et après ça tu viens me flatter ! Pour moi, ça ne marche pas. C'est quoi le but ? Moi je ne comprends pas, donc je deviens méfiante.* [...] *Je me méfie, mais pour me protéger. Pour éviter qu'elle me fasse mal encore une fois*. » (Personne avec un TPL, dyade 9)	« *Bien, moi ce que j'en comprenais, c'est que c'est une dame qui se disait très fragile aux commentaires des autres, qui pouvait être très réactive aux commentaires des autres et qui va souvent les percevoir comme des attaques ou des rejets ou des critiques sévères à son égard. Puis quand elle perçoit les choses de cette façon-là, elle a d'emblée tendance à prendre une position sur la défensive. Elle devient hypervigilante, commence à être méfiante au niveau relationnel, puis elle commence à voir les choses d'une façon très négative. Donc, peu importe si les propos sont nuancés ou non à son égard, si elle a reçu une critique elle va commencer à tout voir sur un pôle très négatif.* » (Intervenant, dyade 9)

CONCLUSION

L'étude de cas présente un intérêt pour les chercheurs qui veulent l'utiliser dans l'analyse d'un phénomène peu connu, tout en tenant compte de son contexte. L'étude de cas présente une forte validité interne, mais peut comporter certaines limites concernant sa validité externe. Ces limites peuvent toutefois être dépassées par l'application de stratégies qui assurent la rigueur scientifique du processus de recherche (p. ex., étudier plus d'un cas). De plus, il est intéressant de constater que les résultats de l'étude de cas peuvent servir à divers usages : l'élaboration ou la vérification d'une théorie (Noiseux *et al.*, 2010), l'évaluation de l'implantation d'un programme (voir le chapitre 9 de cet ouvrage), l'analyse de besoins (Levasseur *et al.*, 2012) ou la conception d'un outil de mesure. D'ailleurs, dans le cas de la présente illustration, les résultats de l'étude ont été utilisés dans le but de concevoir un outil de mesure portant sur l'évaluation des obstacles et des facilitateurs de la participation au travail de personnes présentant un TPL, en considérant leur situation socioprofessionnelle particulière. Pour finir, l'étude de cas est une approche de recherche rigoureuse qui offre une compréhension exhaustive, unique et holistique de l'objet d'étude dans son contexte naturel.

RÉFÉRENCES

American Psychiatric Association – APA (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4e éd., Washington, American Psychiatric Press.

Benbasat, I., D.K. Goldstein et M. Mead (1987). « The case research strategy in studies of information systems », *MIS Quarterly*, vol. 11, no 3, p. 369-386.

Corbière, M. et M.-J. Durand (dir.) (2011). *Du trouble mental à l'incapacité au travail. Une perspective transdisciplinaire qui vise à mieux saisir cette problématique et à offrir des pistes d'intervention*, Québec, Presses de l'Université du Québec.

Cramer, V., S. Torgersenet et E. Kringlen (2006). « Personality disorders and quality of life. A population study », *Comprehensive Psychiatry*, vol. 47, no 3, p. 178-184.

Creswell, J.W., W.E. Hanson, V.L. Clark Plano et A. Morales (2007). « Qualitative research designs : Selection and implementation », *The Counseling Psychologist*, vol. 35, no 2, p. 236-264.

Fortin, M.F. (dir.) (2010). *Fondements et étapes du processus de recherche. Méthodes quantitatives et qualitatives*, 2e éd., Montréal, Chenelière éducation.

Gagnon, Y. (2012). *L'étude de cas comme méthode de recherche*, 2e éd., Québec, Presses de l'Université du Québec.

Ghesquière, P., B. Maes et R. Vandenberghe (2004). « The usefulness of qualitative case studies in research on special needs education », *International Journal of Disability, Development & Education*, vol. 51, no 2, p. 171-184.

GUNDERSON, J.G. et P.S. LINKS (2008). *Borderline Personality Disorder : A Clinical Guide*, 2e éd., Washington, American Psychiatric Publishing.

HANLEY-MAXWELL, C., I.A. HANO et M. SKIVINGTON (2007). « Qualitative research in rehabilitation counseling », *Rehabilitation Counseling Bulletin*, vol. 50, no 2, p. 99-110.

HENTZ, P. (2012). « Case study : The method », dans P.L. Munhall (dir.), *Nursing Research : A Qualitative Perspective*, 5e éd., Sudbury, Jones & Bartlett Learning, p. 359-369.

LARIVIÈRE, N., J. DESROSIERS, M. TOUSIGNANT et R. BOYER (2010). « Exploring social participation of people with cluster B personality disorders », *Occupational Therapy in Mental Health*, vol. 26, no 4, p. 375-386.

LEVASSEUR, M., N. LARIVIÈRE, N. ROYER, J. DESROSIERS, P. LANDREVILLE, P. VOYER, N. CHAMPOUX, H. CARBONNEAU et A. SÉVIGNY (2012). « Concordance entre besoins et interventions de participation des aînés recevant des services d'aide à domicile », *Gérontologie et Société*, vol. 143, p. 111-131.

LUCK, L., D. JACKSON et K. USHER (2006). « Case study : A bridge across the paradigms », *Nursing Inquiry*, vol. 13, no 2, p. 103-109.

MARTINSON, K. et C. O'BRIEN (2010). « Conducting case studies », dans J.S. Wholey, H.P. Hatry et K.E. Newcomer (dir.), *Handbook of Practical Program Evaluation*, 3e éd., San Francisco, Jossey-Bass, p. 163-181.

MERRIAM, S.B. (1988). *Case Study Research in Education : A Qualitative Approach*, San Francisco, Jossey-Bass.

MILES, M.B. et A.M. HUBERMAN (2003). *Analyse des données qualitatives*, Paris, De Boeck Université.

NOISEUX, S., D. TRIBBLE ST-CYR, E. CORIN, P. ST-HILAIRE, R. MORISSETTE, C. LECLERC *et al.* (2010). « The process of recovery of people with mental illness : The perspectives of patients, family members and care providers : Part 1 », *BMC Health Services Research*, vol. 10, p. 161-161.

ROSENBERG, J.P. et P.M. YATES (2007). « Schematic representation of case study research designs », *Journal of Advanced Nursing*, vol. 60, no 4, p. 447-452.

SKODOL, A.E., J.G. GUNDERSON, T.H. MCGLASHAN, I.R. DYCK, R.L. STOUT, D.S. BENDER *et al.* (2002). « Functional impairment in patients with schizotypal, borderline, avoidant, or obsessive-compulsive personality disorder », *American Journal of Psychiatry*, vol. 159, no 2, p. 276-283.

SORIN-PETERS, R. (2004). « The case for qualitative case study methodology in aphasia : An introduction », *Aphasiology*, vol. 18, no 10, p. 937-949.

STAKE, R.E. (1994). « Case studies », dans N.K. Denzin et Y.S. Lincoln (dir.), *Handbook of Qualitative Research*, Thousand Oaks, Sage Publications, p. 236-247.

STAKE, R.E. (1995). *The Art of Case Study Research*, Thousand Oaks, Sage Publications.

WALSHE, C.E., A.L. CARESS, C. CHEW-GRAHAM et C.J. TODD (2004). « Case studies : A research strategy appropriate for palliative care ? », *Palliative Medicine*, vol. 18, no 8, p. 677-684.

YIN, R.K. (1981). « The case study crisis : Some answers », *Administrative Science Quarterly*, vol. 26, nº 1, p. 58-65.

YIN, R.K. (1999). « Enhancing the quality of case studies in health services research », *Health Services Research*, vol. 34 (5 part II), p. 1209-1224.

YIN, R.K. (2009). *Case Study Research : Design and Methods*, 4e éd., Thousand Oaks, Sage Publications.

CHAPITRE 5

LA THÉORISATION ANCRÉE
Une théorisation ancrée pour l'étude de la transition des perceptions de l'état de santé

Marie-Claude Jacques
Maude Hébert
Frances Gallagher
Denise St-Cyr Tribble

FORCES

- Elle comporte des étapes d'analyse détaillées qui contribuent à sa rigueur.
- Elle vise à examiner en profondeur les processus, ce qui permet de mettre en lumière les relations entre les problématiques et les structures sociales.
- La théorie générée procure un langage commun qui aide les acteurs intéressés à mieux comprendre un phénomène et à permettre la mise en œuvre d'actions.

LIMITES

- Elle est fondée sur plusieurs notions complexes ou abstraites qui peuvent la rendre difficile à maîtriser.
- Une analyse incomplète mènera à une description qui ne peut être qualifiée de théorisation ancrée.
- Il peut être difficile pour le chercheur, au moment de l'analyse des données, de mettre en suspens ses préconceptions et sa perspective disciplinaire.

La théorisation ancrée est la méthode qualitative la plus répandue des méthodes qualitatives, et ce, dans de nombreuses disciplines et avec des thèmes de recherche variés (Bryant et Charmaz, 2007). Bien qu'elle puisse sembler au premier abord complexe, il faut plutôt y voir le souci que ses différents concepteurs ont eu, au fil des années, de fournir aux chercheurs qualitatifs une méthode rigoureusement détaillée, mais composée de techniques et de procédures se voulant accessibles.

Ce chapitre est divisé en deux principales parties. Dans la première partie, les éléments essentiels à l'utilisation judicieuse de la théorisation ancrée sont présentés, suivis de ses principales procédures et étapes. En deuxième partie, la méthode est illustrée explicitement à l'aide de la présentation d'une étude portant sur le processus de transition des perceptions de l'état de santé de femmes atteintes d'un cancer du sein.

1. PRÉSENTATION DE LA MÉTHODE

La théorisation ancrée (ou théorisation enracinée) est une méthode inductive développée par Glaser et Strauss (1967) qui a pour but de construire une théorie à partir des données, selon une approche où les données sont simultanément recueillies et analysées. La méthode proposée par Glaser et Strauss, dans sa première version, se voulait rattachée au courant positiviste, en offrant une base solide, systématique et claire pour l'analyse des données et la construction de théories (Bryant et Charmaz, 2007). Au fil des années, Glaser et Strauss ont pris des chemins divergents quant à leur interprétation de la théorisation ancrée. En effet, leur position initiale soutenait que la réalité est indépendante du chercheur et des méthodes employées pour l'étudier, et que l'on peut construire une théorie suffisamment générale pour être applicable à diverses situations et populations (Glaser et Strauss, 1967).

Par la suite, Glaser est resté davantage fidèle à cette position classique positiviste (Glaser, 2007), alors que Strauss, dans ses travaux avec Juliet Corbin, s'est montré peu à peu plus flexible à ce sujet, tourné vers un paradigme plus post-positiviste, mettant de l'avant que la réalité est complexe, située dans un contexte et qu'elle ne peut être cernée dans sa totalité. Ainsi, le chercheur ne peut prétendre arriver à représenter une réalité de façon totalement objective, mais vise plutôt l'interprétation la plus plausible des données observées (Corbin et Strauss, 2008). Charmaz (2006), de son côté, développera au fil des années une théorisation ancrée qu'elle considère comme plus constructiviste. De ce dernier point de vue, la réalité est construite à partir des interprétations du participant, qui influenceront ensuite le chercheur, qui fera ses propres interprétations de cette réalité. Ainsi, « l'interprétation du phénomène étudié est elle-même une

construction» (Charmaz, 2006, p. 187; traduction libre). Cela dit, toutes ces approches de la théorisation ancrée coexistent de nos jours et, comme le suggèrent Heath et Cowley (2004), les chercheurs novices devraient choisir une approche qui correspond à leur style cognitif et qui les aidera à développer leurs habiletés d'analyse. Le tableau 5.1 résume les distinctions entre les approches de Glaser, de Corbin et Strauss et de Charmaz.

Tableau 5.1.
Distinctions entre trois approches courantes de la théorisation ancrée

Type	**Classique** (Glaser et Strauss, 1967 ; Glaser, 1992, 2007)	**Straussienne** (Strauss et Corbin, 1998 ; Corbin et Strauss, 2008)	**Constructiviste** (Charmaz, 2000, 2005, 2006)
Paradigme	Positiviste/objectiviste	Post-positiviste/pragmatique	Constructiviste
Épistémologie	La réalité est considérée comme totalement objective et externe.	La réalité peut être découverte par la recherche, mais ne peut jamais être entièrement appréhendée.	La réalité est une construction issue des interprétations.
Relation avec les participants	Indépendante	Active	Coconstruction
Phases de l'analyse de données	Codage ouvert Codage sélectif Codage théorique	Codage ouvert Codage axial Codage sélectif	Codage initial Codage sélectif Codage théorique

Sources : Birks et Mills, 2011 ; Charmaz, 2006 ; Corbin et Strauss, 2008 ; Ghezeljeh et Emami, 2009 ; Hall, Griffiths et McKenna, 2013 ; Hunter *et al.*, 2011 ; Tan, 2010.

1.1. Utilisation adéquate de la théorisation ancrée

On parle réellement de théorisation ancrée lorsque les résultats de recherche constituent un processus composé de concepts interreliés avec un degré d'abstraction élevé, qui ont émergé de données brutes (Birks et Mills, 2011). La théorie est présentée sous forme de concepts interreliés et il ne s'agit donc pas d'une liste de thèmes. Plus précisément, il faut s'assurer que la recherche a permis de répondre à une question portant sur un processus, plutôt que sur une description ou une interprétation d'un phénomène (voir le chapitre 1 de cet ouvrage). La recherche devrait avoir pour visée l'émergence d'une théorie qui répond à la question de recherche initiale en illustrant la dynamique du phénomène d'intérêt. Ensuite, la boucle répétée «analyse des données-rédaction de mémos-échantillonnage théorique-analyse des données…» est essentielle à la théorisation. En effet, on souhaite dégager

les catégories principales, ainsi que les relations entre celles-ci. La comparaison constante des données aux catégories théoriques fait aussi partie des conditions essentielles pour se distinguer du modèle qualitatif générique (Hood, 2007).

1.2. Objectifs de la méthode et questions de recherche

Au-delà de la description de phénomènes (réponse à la question « Quoi ? »), la théorisation ancrée permet de répondre à la question « Comment ? ». Elle se centre sur le processus en œuvre exprimé par une catégorie centrale autour de laquelle ce processus s'articule (Hood, 2007). Elle est indiquée pour l'étude de phénomènes peu connus ou dont les perspectives théoriques connues sont insatisfaisantes ou incomplètes (Wuest, 2012). Plus précisément, cette méthode est centrée sur les processus sociaux, incluant les actions et les interactions humaines. Elle aide particulièrement à mieux comprendre les problèmes sociaux ou les situations auxquelles les personnes doivent s'adapter (Cooney, 2010).

Cette méthode permet donc de mieux expliquer des phénomènes tels que le changement, le processus ou la trajectoire vécue quant à une situation donnée. Voici quelques exemples de questions de recherche en théorisation ancrée : « Comment les parents développent-ils leurs habiletés à éduquer leur enfant atteint de TDAH ? », « Quel est le processus d'adaptation vécu par des jeunes ayant reçu un diagnostic de schizophrénie ? », « Quelle est la trajectoire vécue par les personnes âgées qui ont développé une dépendance aux psychotropes ? ».

Comme il a été mentionné précédemment, l'objectif ultime de la théorisation ancrée est l'élaboration d'une théorie émergeant de l'analyse de données, qui servira à expliquer le phénomène à l'étude. Essentiellement, la théorie peut être soit substantive, soit formelle. Une théorie substantive est une théorie qui s'applique à un domaine particulier, par exemple au processus de dévoilement d'un diagnostic (domaine décisionnel particulier) ou au fait de devenir un jeune de la rue (identité particulière). La théorie formelle est beaucoup plus abstraite et peut donc s'appliquer à un large éventail de domaines empiriques et est en pratique intemporelle (Glaser et Strauss, 1967 ; Glaser, 2007). Il pourrait s'agir du processus décisionnel (peu importe quelle décision ou quel contexte) ou encore de la construction de l'identité (peu importe quelle identité).

1.3. Notions importantes

La méthodologie de la théorisation ancrée peut sembler complexe et abstraite pour le néophyte. Comment peut-on générer une théorie à partir d'une analyse des données ? Une meilleure compréhension de certains concepts essentiels aide à mieux saisir les fondements de cette méthode.

Tout d'abord, la méthodologie repose notamment sur l'*interactionnisme symbolique* et le pragmatisme. L'interactionnisme symbolique est une perspective développée par Blumer (1969), centrée sur les relations dynamiques entre les significations et les actions. Les significations sont générées par les actions, qui à leur tour influencent les significations. Par exemple, si pour une personne avoir une maladie mentale signifie être enfermée dans une salle capitonnée avec une camisole de force (signification 1), elle pourrait prendre toutes les mesures nécessaires pour masquer ses symptômes de psychose (action 1). Aussi, elle pourrait secrètement faire des recherches sur Internet (action 2) et lire un témoignage d'une personne ayant eu une psychose qui a bénéficié d'un traitement, lequel lui a permis de retrouver une vie satisfaisante. Cette action peut générer une nouvelle signification (2), par exemple que la psychose est une maladie qui peut être soignée. Elle peut alors décider de demander de l'aide (action 3). Cette boucle significations-interactions constitue un processus qui peut être mis en relief en s'intéressant aux perceptions des personnes vivant un phénomène donné, ce que la méthodologie de la théorisation ancrée permet d'ailleurs de réaliser.

Par ailleurs, la théorisation ancrée vise la compréhension d'un *processus*. Un processus est composé de séquences d'actions-interactions évolutives. La recherche de processus dans les données empiriques va influencer la façon de les analyser. Plus précisément, plutôt que de rechercher uniquement des dimensions et des propriétés, l'analyse visera aussi à mettre en lumière les actions et les interactions qui ont évolué dans le temps. Elle servira aussi à déterminer comment ces changements sont survenus, ou ce qui leur permet de rester inchangés tout en ayant une évolution de leurs conditions structurelles (Strauss et Corbin, 1998).

Il est important de noter que dans la méthode de théorisation ancrée, la collecte et l'analyse des données s'effectuent simultanément. Le détail de la technique d'analyse sera abordé plus loin, mais mentionnons, pour le moment, qu'à la suite de l'analyse des premières entrevues, la trajectoire générale (ou processus) commence à se dégager et ceci permet alors de faire un *échantillonnage ciblé*, c'est-à-dire des choix raisonnés quant aux prochains participants à recruter (Richards et Morse, 2007). Par la suite, les catégories émergentes (générées par l'analyse des données), qui permettent au chercheur de comprendre plus profondément la théorie en développement,

vont orienter le recrutement de nouveaux participants à l'étude vers des critères beaucoup plus précis concernant les réponses particulières à une expérience, ou à un concept particulier qui semble significatif. Cette façon de procéder se nomme l'*échantillonnage théorique* (Charmaz, 2006; Richards et Morse, 2007). Il peut aussi s'agir de retourner auprès de participants déjà interviewés et qui ont une expérience d'intérêt par rapport aux catégories émergentes, ou encore de se tourner vers d'autres sources de données comme des rapports, des vidéos ou des écrits existants sur un concept à l'étude (Charmaz, 2006). La *sensibilité théorique* du chercheur est nécessaire pour une bonne analyse de ce processus. Il s'agit de l'habileté à générer des concepts à partir de données et de les relier entre eux. Ceci nécessite que le chercheur soit capable de maintenir une distance par rapport à son analyse, de considérer de multiples points de vue, de voir les possibilités, etc. Il doit aussi être en mesure de développer des perspectives théoriques et de concevoir des idées conceptuelles abstraites (Charmaz, 2006; Glaser, 2007).

Par conséquent, la collecte des données se termine non pas par une saturation des données, mais par une *saturation théorique*, c'est-à-dire lorsque les diverses catégories conceptuelles sont complètement élaborées et que de nouvelles données n'apportent pas ou peu de nouvelles informations utiles à la conceptualisation, bien que de subtiles variations puissent toujours être découvertes (Corbin et Strauss, 2008).

La finalité d'une théorisation ancrée est la production d'une théorie. Pour Corbin et Strauss (2008), une *théorie* est un ensemble de catégories bien développées qui sont systématiquement interdépendantes par l'entremise des énoncés de relation, pour expliquer un phénomène donné. C'est donc par la formulation de liens entre les catégories que la théorisation se distingue de la simple description. La théorie procure un langage commun qui permet aux participants à la recherche, aux professionnels et autres acteurs intéressés de mieux comprendre un problème, de le voir de manière plus générale et ensuite d'entreprendre des actions pour y remédier (Strauss et Corbin, 1998).

1.4. Procédures et étapes

Avant d'aborder les différentes étapes de la méthode d'analyse, la question de la *perspective disciplinaire* sera abordée. Partant du principe que le chercheur ne peut être complètement neutre, Corbin et Strauss (2008) mentionnent qu'il faut en effet tenir compte de la perspective disciplinaire, qui va influencer l'interprétation des données. Pour Charmaz (2006), la discipline doit être un point de départ et non une finalité. Il faut donc, dès les premières phases de l'analyse des données, être le plus ouvert possible à

tout élément observé ou ressenti en laissant de côté sa discipline de prédilection. Les prochaines lignes abordent la conceptualisation et la réalisation d'une recherche utilisant la théorisation ancrée, et ce, sous le format de sept étapes : 1) formulation de la question ; 2) recension des écrits scientifiques ; 3) choix des sources de données ; 4) échantillonnage et recrutement ; 5) analyse des données ; 6) phases de l'analyse des données ; et 7) validation du schème théorique.

1.4.1. Formulation de la question

Comme décrit précédemment, la question de recherche doit s'intéresser à un processus.

1.4.2. Recension des écrits

La recension des écrits n'est pas obligatoirement exclue en théorisation ancrée. Bien qu'à l'origine Glaser et Strauss (1967) aient insisté sur l'importance de commencer la recherche avec un esprit ouvert et neutre, plus tard, Strauss et Corbin (1998) déclareront qu'il est impossible d'être complètement neutre et que la recension des écrits peut, entre autres, permettre d'énoncer certaines questions pour la collecte initiale des données. Elle peut aussi servir de source de données pour l'échantillonnage théorique et éviter de reproduire ce qui a déjà été bien documenté (Charmaz, 2006 ; Corbin et Strauss, 2008 ; Labelle, Navarro-Flores et Pasquero, 2013).

1.4.3. Choix des sources de données

Pour Morse (2007, p. 231 ; traduction libre), « un excellent participant pour une théorisation ancrée est celui qui a vécu, ou observé, le phénomène à l'étude ». Les participants sont souvent la source initiale des données. Au fil de l'analyse des données et de l'échantillonnage théorique, d'autres sources de données pourront être utilisées, incluant les notes de terrain et les écrits déjà existants.

1.4.4. Échantillonnage et recrutement

Il est suggéré de débuter avec un échantillon dit « de convenance », donc déterminé à partir de critères initiaux, de façon à obtenir une vue d'ensemble du phénomène. L'échantillonnage deviendra ensuite un échantillonnage dit « théorique ». Il est donc difficile d'établir précisément, avant d'avoir commencé la recherche, combien de participants seront nécessaires. En effet, de nouvelles données seront recueillies jusqu'à la saturation théorique,

difficile à prédire, car elle sera basée sur l'orientation que prendra l'analyse des données. Cependant, les évaluations requises par les organismes subventionnaires ou les comités d'éthique de la recherche exigent un nombre approximatif de participants qui seront recrutés; il faut pouvoir leur procurer une estimation du nombre de participants requis. Morse (1994) suggère que 30 à 50 entrevues sont nécessaires dans une étude utilisant la théorisation ancrée, tandis que, plus récemment, Wuest (2012) mentionne qu'une théorisation ancrée effectuée dans un domaine bien délimité peut être menée à terme avec des entrevues auprès de 10 à 15 participants. Morse (2007) précise en outre que plus le contenu des entrevues est bien ciblé, meilleure est la qualité des données; donc, moins d'entrevues sont nécessaires et ainsi, moins de participants doivent être recrutés. En bref, le nombre de participants peut être estimé selon les caractéristiques du protocole de recherche.

1.4.5. Analyse des données

Les données doivent commencer à être analysées dès la première entrevue terminée. Suite aux premières entrevues analysées, l'échantillonnage devient théorique, c'est-à-dire fondé sur les caractéristiques du processus général qui émerge des données (Corbin et Strauss, 2008; Morse, 2007). La collecte et l'analyse des données sont donc simultanées. L'analyse comporte diverses phases, ainsi que l'utilisation de mémos. Le détail des phases est expliqué plus loin, cependant, il importe de retenir que l'analyse génère dès le début des concepts et des catégories. Les *concepts* sont des mots représentant les idées générées par les données (p. ex., affection). Les *catégories* sont des concepts de haut niveau qui chapeautent un ensemble de concepts de niveau inférieur partageant les mêmes *propriétés* (caractéristiques qui définissent et décrivent le concept). Par exemple, «affection», «encouragement» et «conseil» pourraient être dans la catégorie «soutien social». La rédaction de *mémos* tout au long de l'analyse des données est une étape cruciale pour la théorisation ancrée. Différents des notes d'observation ou d'autres descriptions réalisées en marge de la collecte des données, les mémos sont un outil d'analyse. Leur rédaction est consacrée à la conceptualisation, ainsi qu'aux relations entre les données, les concepts et les catégories. Les mémos sont essentiels pour atteindre le niveau d'abstraction nécessaire à la théorisation (Bryant et Charmaz, 2007). L'utilisation de *diagrammes* est aussi fortement recommandée par plusieurs auteurs pour l'analyse des données (Charmaz, 2006; Corbin et Strauss, 2008; Glaser, 2007). Les diagrammes sont des représentations visuelles (p. ex., une carte conceptuelle) qui permettent d'illustrer les relations entre les concepts en cours d'analyse. Un diagramme peut aussi servir à la représentation finale de la théorie générée lorsque l'analyse est terminée.

1.4.6. Phases de l'analyse des données

L'examen des diverses méthodes de la théorisation ancrée permet de constater que les auteurs ont des façons différentes de nommer les niveaux de codage (voir le tableau 5.1). Toutefois, peu importe l'approche choisie, il y a trois phases de codage principales qui peuvent être qualifiées d'initiale, d'intermédiaire et d'avancée (Birks et Mills, 2011). Dans le présent chapitre, nous présenterons la nomenclature utilisée par Strauss et Corbin (1998), qui est celle utilisée dans l'application de la théorisation ancrée présentée en deuxième partie du chapitre.

Codage ouvert

Il s'agit d'abord de faire une lecture approfondie des données recueillies pour en faire une analyse ligne par ligne, ou encore paragraphe par paragraphe, qui permettra de générer des concepts. Il est normal que des dizaines de concepts émergent de cette procédure. Or il apparaîtra que certains de ces concepts peuvent être regroupés dans des concepts abstraits de plus haut niveau qui décrivent des actions. Strauss et Corbin (1998) donnent l'exemple des concepts « oiseau », « cerf-volant » et « avion », qui ont le point commun de « voler ». Le concept de niveau plus élevé est une catégorie aux propriétés (p. ex., se déplace dans les airs) et aux dimensions (p. ex., se déplace dans les airs de lentement à rapidement) déterminées. Les dimensions servent à établir la variabilité d'une catégorie. La *comparaison constante* est un outil d'analyse très important à cette étape-ci. Il s'agit notamment de comparer les incidents aux incidents (tels que présentés par les participants), ou encore les catégories aux catégories, de façon à en discerner les similarités ou les différences qui permettront de les classer.

Codage axial

Ce type de codage consiste à établir les relations des catégories avec leurs sous-catégories, pour une compréhension plus précise et complète du phénomène. Ces relations sont fondées sur les propriétés et les dimensions des catégories et des sous-catégories. Strauss et Corbin (1998) classent ces relations en conditions causales (événement ou circonstance qui influence le phénomène), intermédiaires (atténue ou altère l'effet de la condition causale) ou contextuelles (conditions particulières qui se croisent à un moment et dans un lieu donnés et qui créent l'ensemble des circonstances ou des problèmes sur lesquels la personne doit agir ou interagir).

Codage sélectif

C'est à cette étape que s'effectue la théorisation, où les catégories sont intégrées et raffinées. L'intégration consiste principalement en l'identification d'une catégorie centrale qui rattache les autres catégories entre elles, de façon à former un tout explicatif. Diverses techniques peuvent être utilisées à cette fin, notamment la rédaction de mémos, de l'histoire (écrit qui tente de répondre à ce qui se produit, qui cherche à expliquer le sens général), ou encore les diagrammes. Le raffinement de la théorie s'effectue par la révision de la structure globale du processus (schème). Il s'agit ici de s'assurer de sa cohérence interne (le schème est logique), de saturer les catégories insuffisamment développées (jusqu'à ce que l'analyse n'apporte plus de nouvelles propriétés ou dimensions) et d'élaguer celles qui sont en excès (contribuent peu, ou aucunement, à la compréhension du schème théorique).

1.4.7. Validation du schème théorique

La dernière étape de la théorisation ancrée consiste à valider la théorisation. La validation signifie ici de s'assurer que le schème théorique correspond aux données brutes. Elle permet aussi de s'assurer que rien d'essentiel n'a été oublié. Le schème théorique devrait pouvoir expliquer la grande majorité des cas. Il peut s'agir de retourner aux données brutes, ou encore de revoir les participants, de leur décrire l'histoire représentée par le schème théorique et de leur demander à quel point cela correspond à leur cas (au sens large). Ceci peut se faire de diverses façons : en personne, par la poste, par téléphone ou par courriel.

1.5. Utilisation de logiciels d'analyse en théorisation ancrée

En général, les auteurs perçoivent des avantages à l'utilisation de logiciels pour la recherche qualitative, mais d'autres peuvent y voir des obstacles à la qualité de l'analyse en théorisation ancrée. Corbin et Strauss (2008), par exemple, abordent l'utilité des logiciels dès les premières pages de leur dernier ouvrage. Cependant, il est recommandé que le logiciel soit un soutien au processus d'analyse plutôt que destiné à le diriger. Charmaz (2011), par contre, est plutôt en défaveur d'une telle utilisation et note que les logiciels seraient utiles pour codifier des sujets et des thèmes, mais ne le seraient pas pour codifier des processus et s'engager dans la comparaison constante. Ses réticences sont aussi associées au fait que les possibilités des logiciels seraient incompatibles avec le paradigme constructiviste, car ils rendraient difficile d'effectuer la recherche dans une réelle approche de coconstruction des données et de la théorie avec les participants. Enfin, d'autres auteurs ont précisément documenté l'utilisation d'un logiciel pour

mener à terme une théorisation ancrée. De leur point de vue, l'utilisation de logiciels est concluante pour ce type de méthode. En effet, un logiciel peut être utilisé pour la théorisation ancrée, mais selon certaines conditions très précises et axées prioritairement sur la transparence de la démarche (Bringer, Johnston et Brackenridge, 2004; Hutchison, Johnson et Breckon, 2010).

2. ILLUSTRATION DE LA THÉORISATION ANCRÉE

Cette partie illustre l'utilisation d'un devis de théorisation ancrée dans une étude doctorale portant sur le processus de transition des perceptions de l'état de santé de femmes atteintes d'un cancer du sein (Hébert, Gallagher et St-Cyr Tribble, soumis). Cette présentation fait donc état du contexte de l'étude, des objectifs de recherche, de la collecte et de l'analyse des données. Finalement, quelques résultats et leur interprétation seront présentés.

2.1. Contexte de l'étude

Chaque année, près de 1,38 million de femmes dans le monde reçoivent un diagnostic de cancer du sein, ce qui fait de ce cancer le plus prévalent chez les femmes, tant dans les pays occidentaux que dans les pays en voie de développement (OMS, 2013). À la suite de l'annonce d'un diagnostic de cancer du sein, les femmes reçoivent des informations sur les interventions et les traitements, mais peu d'attention est accordée à leurs perceptions du cancer du sein et de leur état de santé (Hébert et Côté, 2011). En effet, les perceptions des femmes de cette maladie demeurent peu explorées par les infirmières (Park *et al.*, 2009).

Pourtant, l'évaluation de ces perceptions est d'une importance capitale, car elles déterminent le niveau de détresse psychologique chez les femmes diagnostiquées d'un cancer du sein, les stratégies d'adaptation utilisées, ainsi que l'adhésion aux traitements proposés (Leventhal *et al.*, 1997; Meleis, 2010). Si les perceptions de l'état de santé des femmes atteintes d'un cancer du sein étaient mieux comprises de la part des professionnels de la santé, il est possible de croire que les interventions mises en œuvre diminueraient les délais de consultation liés aux attributions bénignes des symptômes. De plus, ces femmes recevraient des soins infirmiers plus humanisés, basés sur une relation d'aide et de confiance, et tenant compte de leurs perceptions tout au long de la trajectoire de la maladie. À l'examen des écrits scientifiques, force est de constater la quasi-inexistence d'études sur le processus de transition de l'état de santé, qui nous permettraient de mieux comprendre ce que les femmes vivent entre être en santé et se retrouver, du

jour au lendemain, diagnostiquées d'une maladie grave telle que le cancer du sein. La théorisation ancrée est alors apparue comme la méthodologie de choix puisqu'elle sert à comprendre en profondeur et à élucider des processus sociaux, tout en développant une théorie fondée empiriquement, et ce, à partir du point de vue des acteurs qui vivent ce processus.

2.2. Objectifs de recherche

Le but de cette recherche était de modéliser le processus de transition des perceptions de l'état de santé de femmes atteintes d'un cancer du sein. Un processus consiste en un flux d'émotions, de réflexions et d'actions (Corbin et Strauss, 2008) par lesquelles sont passées les femmes lors de la transition de leurs perceptions relatives à leur état de santé. Pour ce faire, nous nous sommes posé la question suivante : Comment les femmes vivent-elles le processus de transition des perceptions de l'état de santé après avoir reçu un diagnostic de cancer du sein ? L'approche de Strauss et Corbin (1998) a été retenue, car elle prend en compte la complexité et la variabilité des phénomènes sociaux en plus d'être sensible à la nature évolutive des événements, et ce, tout en reconnaissant l'interrelation entre le contexte, l'action et les conséquences. Par ailleurs, ces auteurs décrivent leur méthode de recherche de façon détaillée, ce qui en facilite l'application.

2.3. Présentation des étapes de la recherche

2.3.1. Participantes à la recherche

Voulant étudier les perceptions de femmes atteintes d'un cancer du sein, nous avons fait le recrutement dans des hôpitaux qui offrent des soins tertiaires et qui possèdent une clinique ambulatoire d'oncologie. Cette décision méthodologique permettait d'augmenter la diversité des situations et ainsi offrir la capacité de renseigner le plus possible sur la transition en contexte réel. Il importait de recueillir des données sur le milieu de soins afin de favoriser la compréhension du processus de transition que les femmes vivent. Ayant une expérience clinique en tant qu'infirmière en oncologie, l'instigatrice de cette étude (Maude Hébert) connaissait bien l'environnement de ces milieux de soins, l'engrenage qui s'enclenche lorsqu'une femme est diagnostiquée d'un cancer du sein et les types de traitements requis. Cela a eu deux bénéfices concrets, soit de permettre à la chercheuse de mieux comprendre ce que les femmes vivent et d'établir un lien de confiance entre elle et les participantes.

L'échantillonnage de convenance a été utilisé pour le choix des premières participantes à l'étude afin d'obtenir une vision d'ensemble du phénomène de la transition. Ainsi, compte tenu de ces principes, les premières participantes retenues étaient des femmes répondant aux critères d'inclusion suivants : 1) être âgée entre 40 et 60 ans ; 2) avoir commencé ou terminé des traitements de chimiothérapie, de radiothérapie ou avoir subi une chirurgie ; 3) ne pas avoir d'antécédents personnels d'autres types de cancer ; 4) ne pas être atteinte d'aphasie ou de troubles psychologiques invalidants ; 5) parler français ; et 6) désirer participer à l'étude. Par la suite, l'échantillon a été continuellement remanié en réponse aux analyses. Ainsi, la chercheuse a interviewé des femmes se situant à divers moments de la trajectoire de la maladie et du traitement, soit des femmes nouvellement diagnostiquées d'un cancer, qui ont : vécu jusqu'à trois récidives ; subi une ablation de la tumeur ; subi l'ablation totale des seins en plus des traitements de chimiothérapie et de radiothérapie ; subi une reconstruction mammaire ; été diagnostiquées au stade 1 de la maladie (maladie précoce) ; ainsi que des femmes diagnostiquées au stade 4 (cancer métastatique).

Par la suite, la chercheuse a recruté des cas extrêmes qui ont servi à cerner les limites du phénomène, tels que des femmes en santé, des femmes en santé, mais porteuses du gène qui les met à risque de développer un cancer du sein et des femmes en soins palliatifs d'un cancer du sein. Ces données recueillies ont permis de dégager les conditions d'apparition et de variation du phénomène pour ainsi favoriser la théorisation de la transition des perceptions de la santé à la maladie (Strauss et Corbin, 1990 ; 1998). Bien qu'il soit difficile de déterminer à l'avance la taille exacte de l'échantillon, nous avions prévu de rencontrer une trentaine de femmes. De ce fait, nous avons rencontré 32 femmes et nous avons remarqué que les quatre dernières entrevues n'apportaient pas de nouveauté à la compréhension du phénomène.

Après l'obtention de l'approbation des différents comités d'éthique de la recherche, nous avons amorcé le recrutement des participantes. Elles ont été recrutées parmi la clientèle de deux centres hospitaliers, d'une communauté universitaire et dans la population de la région où se déroulait l'étude. L'infirmière pivot d'une clinique ambulatoire d'oncologie a facilité le recrutement de femmes en soins palliatifs (centre hospitalier). Des femmes atteintes d'un cancer ont manifesté leur intérêt à participer à la recherche à la suite d'annonces parues dans les journaux locaux, et d'affiches posées dans les cliniques ambulatoires d'oncologie ainsi que dans les salles d'attente de groupes de médecine de famille. Le recrutement s'est aussi déroulé dans une université par l'intermédiaire d'une annonce. Ainsi, les femmes désireuses de participer à l'étude communiquaient avec

la chercheuse, qui leur expliquait la recherche, répondait à leurs interrogations, et, si elles désiraient participer aux entrevues, la chercheuse planifiait un rendez-vous avec elles.

Les femmes interviewées ont été rencontrées dans un lieu de leur choix (leur domicile, le bureau de la chercheuse, un café ou un parc public) et au moment qui leur convenait le mieux. Dans un premier temps, lors de ces rencontres, la chercheuse lisait le formulaire d'information et de consentement avec les participantes, puis elle répondait aux questions posées par ces dernières. Dans un deuxième temps, les femmes répondaient au questionnaire sociodémographique. Puis, dans un troisième temps, l'entrevue de type semi-dirigée, enregistrée en mode audionumérique, d'une durée variant entre 60 et 90 minutes, avait lieu. Chaque participante a été rencontrée une seule fois. La chercheuse a transcrit le contenu de chacune des entrevues.

2.3.2. Collecte des données

Trois outils de collecte des données ont été utilisés afin d'obtenir une compréhension de la transition des perceptions et de modéliser les perspectives des participantes. Il s'agissait d'entrevues individuelles semi-dirigées, des notes de terrain à propos du déroulement des entrevues (p. ex., « La participante semblait mal à l'aise de raconter son vécu lorsque son mari est entré dans la pièce ») et des mémos concernant chaque nouvelle idée méthodologique, et ce, tout au long de la collecte et de l'analyse des données (p. ex., « La création d'une nouvelle catégorie est nécessaire puisque plusieurs femmes ont parlé de l'importance d'être positive pour traverser la maladie »).

Le guide d'entrevue incluait des questions ouvertes avec des questions d'approfondissement. Voici quelques questions posées lors de chaque entrevue : « De quelle façon voyiez-vous le cancer du sein avant d'en être atteinte ? », « De quelle façon voyez-vous votre cancer aujourd'hui ? » et « Quelles sont les étapes que vous avez traversées pour en venir à voir la santé et la maladie telles que vous les voyez aujourd'hui ? ». Des questions d'approfondissement telles que « Que voulez-vous dire par… ? » « Pouvez-vous m'en dire davantage ? » permettaient de mieux saisir le sens de l'idée exprimée par les participantes. Tout au long de l'entrevue, les hypothèses inductives étaient validées avec la participante et les questions évoluaient en même temps que la compréhension du phénomène. En guise d'exemple, la chercheuse a apporté une modification à son guide d'entrevue après qu'une participante lui eut dit ne pas se sentir malade en réponse à la question « Comment en êtes-vous venue à vous sentir malade du cancer du sein ? ». Afin de bonifier sa quête de la compréhension du point de vue

des participantes, la chercheuse a modulé ainsi ses questions : « Qu'est-ce qu'une maladie pour vous ? » « Qu'est-ce qui fait en sorte que vous ne vous sentez pas malade ? »

2.4. Analyse des données basée sur la comparaison constante

L'analyse des données s'est déroulée en concomitance avec la collecte des données, toujours pour respecter le principe de circularité de la théorisation ancrée (Charmaz, 2006 ; Corbin et Strauss, 2008 ; Luckerhoff et Guillemette, 2013 ; Strauss et Corbin, 1998). Cela s'illustre par les nombreux allers et retours entre les données, l'analyse et les écrits faits lors de discussions en équipe ou de la lecture des mémos. Ces discussions, riches de sens, généraient de nouvelles idées ou hypothèses, qui entraînaient des modifications du guide d'entrevue dans le but d'étayer la théorie en développement. Par exemple, la catégorie « être positive » était placée au départ sous la rubrique « prise de conscience à la suite de l'annonce du diagnostic ». Après que l'entrevue ait été codée par les trois membres de l'équipe de recherche, il est ressorti que la catégorie « être positive » devait plutôt être placée sous la rubrique du « changement de la conception de la vie » par les femmes pour passer à travers l'épreuve du cancer.

La technique de comparaison constante a été effectuée pendant l'analyse des données, c'est-à-dire que les données émergentes ont été comparées avec les données amassées lors des entrevues précédentes, avec les mémos de la chercheuse et avec des concepts décrits dans la littérature scientifique, tels que la transition, l'adaptation et la croissance personnelle post-traumatique afin de délimiter le phénomène émergent. Cette méthode comparative est au cœur de l'analyse dans la théorisation ancrée. Pour ce faire, la chercheuse a fait ressortir les similitudes et les contrastes entre les données recueillies par les entrevues, la recension des écrits et les mémos. Pour illustrer cette technique, nous avons comparé le déclencheur de la transition entre se sentir en santé et se sentir malade chez les femmes interviewées avec le déclencheur de la transition identifié dans la théorie de la transition développée par Meleis, Sawyer, Im, Messias et Schumacher (2000). Ainsi, cela nous a permis de déterminer que la transition débute à l'annonce du diagnostic de cancer du sein plutôt que dans l'attente de résultats d'examens non concluants (aussi appelés douteux) nécessitant des examens complémentaires. La comparaison constante des données a permis de cerner les caractéristiques et les relations entre les composantes du processus de transition à l'étude (Paillé, 1994 ; Strauss et Corbin, 1998). Ainsi, à la fin des cinq dernières entrevues, la chercheuse demandait l'opinion des participantes

sur la modélisation des résultats qui avait émergé de l'analyse des données. Ceci améliorait, par le fait même, la crédibilité des résultats (Charmaz, 2006 ; Laperrière, 1997 ; Strauss et Corbin, 1998).

Le contenu des entrevues a été transcrit par la chercheuse. Comme décrit précédemment, la codification selon Strauss et Corbin (1998) comprend trois niveaux. Ceux-ci seront illustrés dans les lignes qui suivent. Il est important de revenir sur le fait que la théorisation ancrée est un processus de recherche circulaire qui nécessite une alternance entre les trois niveaux de codage.

2.4.1. Codage ouvert

Cette première étape du codage sert à attribuer un mot (concept) à l'idée principale qui se dégage d'une phrase ou d'un paragraphe. Nous avons lu chaque entrevue en interrogeant les données comme suit : *De quoi est-il question ici ? En face de quel phénomène suis-je en présence ?* (Paillé, 1994 ; Strauss et Corbin, 1998). Nous avons ensuite placé un code sous forme de concept qui résume l'idée principale de la phrase (tableau 5.2).

À titre d'exemple, posons ces questions à un extrait d'entrevue dans lequel une des participantes raconte comment elle s'est sentie lors de l'annonce du diagnostic de cancer. Dans la colonne de droite se trouve le code attribué à chaque extrait du compte rendu intégral (*verbatim*).

Tableau 5.2.
Extraits de l'entrevue réalisée auprès d'une participante (P-1)

Extrait	Code
Bien un choc, complètement atterrée. Là tu es sans mot. La vie arrête complètement. Tout ce qu'il y a autour, c'est comme un film qui passe devant toi puis tu n'es plus dedans du tout.	Choc
J'étais très fâchée. Je les haïssais toutes les filles qui avaient de beaux seins. Moi, je n'ai pas des gros seins : Je me disais : « Pourquoi ? » J'avais de la révolte.	Colère
Une chance que j'avais du soutien de ma famille et de mes amies pour passer à travers cette expérience. Ce qui fait la différence, c'est dans l'approche. Être présent à l'autre. Et le soutien.	Soutien des proches
J'ai essayé beaucoup avec l'humour aussi. J'ai essayé d'en rire le plus possible sinon ça ne donne rien.	Humour
J'ai eu une vie assez difficile. Dépressive. Ça a été dur. Des dépressions c'est plus dur à vivre qu'un cancer du sein à mes yeux à moi.	Comparaison du cancer avec les problèmes de santé mentale

Ici encore, on utilise la comparaison constante pour préciser les propriétés et les dimensions des concepts et des catégories. À titre d'exemple, le concept de choc se définit comme une forte réaction émotive qui survient particulièrement à l'annonce du diagnostic (caractéristiques) et il peut y avoir des nuances de « choc » selon les participantes (dimensions) en termes de durée ou d'intensité, par exemple.

2.4.2. Codage axial

Cette deuxième étape du codage sert à faire des liens entre les différents concepts nommés lors du codage ouvert et à élever le niveau d'analyse à un niveau dit conceptuel qui englobe les phénomènes de façon à former des catégories de concepts. Ces catégories sont formées de concepts ayant des propriétés communes (Paillé, 1994 ; Strauss et Corbin, 1998). On remarque, dans l'exemple précédent, que les concepts « colère » et « choc » peuvent être regroupés sous une même catégorie englobante, soit « les réactions à la suite de l'annonce du diagnostic », et que la catégorie « changement de la conception de la vie » englobe la catégorie « mise en œuvre de stratégies d'adaptation » qui chapeaute, à son tour, les concepts « recherche et acceptation du soutien », « humour » et « comparaison du cancer avec les problèmes de santé mentale ». Ces exemples illustrent bien la distinction entre les concepts, qui résument ou thématisent l'idée principale, et les catégories qui élèvent le niveau de compréhension du phénomène à un contexte plus explicatif et plus large en incluant plusieurs concepts (Paillé, 1994).

À cette étape, la chercheuse s'est posée des questions telles que « Est-ce que ces concepts sont liés ? » « En quoi et comment sont-ils liés ? ». Il peut arriver que certains codes semblent contradictoires. Par exemple, dans le premier extrait du tableau 5.2, la participante témoigne avoir ressenti un choc lors de l'annonce du diagnostic. En comparant ces données avec le *verbatim* d'une autre participante, qui raconte comment elle s'est sentie lors de l'annonce du diagnostic du cancer du sein, on se rend compte que leur expérience est très différente, comme en témoigne cet extrait : « *Je pense que le monde alentour de moi était plus énervé que moi.* [Elle se demandait :] *Est-ce que je suis une extraterrestre moi ? On dirait que je ne suis pas énervée par ça* » (P-25). À première vue, cette différence observée entre les deux discours à propos du même événement, en l'occurrence l'annonce du diagnostic d'un cancer du sein, peut laisser le chercheur perplexe, alors qu'en fait, cela le pousse à différencier les réacions, à mieux comprendre le phénomène d'intérêt et à établir les dimensions du « choc » à l'annonce du diagnostic de cancer du sein. Par exemple, nous avons remarqué que le choc à la suite de l'annonce d'un diagnostic de cancer du sein peut varier de minime à maximal selon le stade du cancer et des représentations que les femmes ont

du cancer. En observant *qui* vit l'annonce du diagnostic comme un choc, *quand*, *comment*, *pourquoi* et *avec quelles conséquences*, le chercheur peut ainsi mieux détailler le contenu de la catégorie et faire des liens avec les autres catégories (Strauss et Corbin, 1990). L'analyse des données a permis aux chercheuses de constater que les femmes qui vivent un plus grand choc au moment du diagnostic (qui) perçoivent le cancer de manière plus grave (cause) et modifient davantage leurs perceptions de la vie (conséquences), comme l'illustre l'extrait suivant : « *Aujourd'hui, je mets des efforts, énormément d'efforts pour ne pas avoir une récidive de cancer. C'est difficile un cancer. Je ne savais pas que c'était difficile comme ça* » (P-3).

Afin de faciliter la mise en relation et la schématisation, les auteurs sur la théorisation ancrée s'entendent sur l'importance de rédiger des mémos (Charmaz, 2006 ; Corbin et Strauss, 2008 ; Glaser et Strauss, 1967 ; Strauss et Corbin, 1998). Voici un exemple de mémo en lien avec l'extrait d'entrevue précédent : « J'ai observé que les femmes semblent vivre le cancer comme un deuil. Passent-elles à travers les mêmes étapes ? Est-ce que certains codes doivent être jumelés ? » En réponse à cette question, nous avons remarqué, entre autres, que certaines dimensions de la transition vécue par les femmes s'apparentent aux étapes du deuil. Nous avons regroupé les concepts où les femmes se demandaient « Pourquoi moi ? » sous le code « colère » avec le non-verbal à l'appui (c.-à-d. l'intonation) de la catégorie « étapes du deuil ». Ces catégories sont chapeautées par la catégorie « les réactions à la suite de l'annonce du diagnostic de cancer du sein ».

Les schémas représentent un autre moyen de favoriser la mise en relation. La figure 5.1 est un exemple de schématisation du processus de transition des perceptions de l'état de santé de femmes atteintes d'un cancer du sein. Cette étape est importante puisqu'elle rend les données dynamiques dans l'explication du phénomène à l'étude (Paillé, 1994). Strauss et Corbin (1998) recommandent de faire une esquisse à la fin de chaque entrevue pour faire ressortir les relations entre les concepts. Cette schématisation simple est réalisée après chaque rencontre avec une participante pour résumer les concepts et les situer dans le temps. Par exemple, puisque la définition de la transition est un changement interne situé entre deux périodes relativement stables, nous avons remarqué que la transition de l'état de santé des femmes atteintes d'un cancer du sein se situe entre l'annonce du diagnostic et la fin des traitements de chimiothérapie et de radiothérapie.

Figure 5.1.
Schématisation des données d'une entrevue

Perceptions de l'état de santé

Avant l'annonce du diagnostic
- Sentiment d'invulnérabilité
- Santé est acquise
- Santé est opposée à la maladie

Diagnostic

Traitements
- Être malade pendant les traitements
- Rassurance contre la récidive
- Âge est un facteur
- Peur de l'opération
- Modification de l'image corporelle : sein et perte de cheveux
- Diminution de la libido
- Choc
- Peur de mourir
- Peine
- Colère
- Injustice
- Pourquoi moi?
- Recherche de cause/échec
- Cancer = monstre

Fin des traitements

Après la transition
- Peur de la récidive : épée de Damoclès
- Santé est un cadeau
- Cycle de deux ans

Prises de conscience
- Cadeau : ça nous rend meilleur
- Changement de priorités : Penser plus à soi
- Se comparer à soi et aux autres
- Rationnaliser pour se consoler
- Développer de l'empathie
- Engagement social = promouvoir le dépistage, aider les autres personnes atteintes
- Ne pas se sentir malade du cancer du sein

Facteurs facilitant la transition
- S'accrocher aux petits-enfants
- Soutien

Facteurs entravant la transition
- Lourdeur du système de santé

2.4.3. Codage sélectif

Cette troisième étape du codage sert à délimiter l'étude qui prend forme et à intégrer les différentes catégories. Le chercheur doit identifier le noyau théorique, nommé catégorie centrale. Cette catégorie résume l'idée centrale de l'étude et peut être comparée au titre d'un film (Glaser et Strauss, 1967; Strauss et Corbin, 1998). Dans la présente recherche, nous avons remarqué qu'un thème récurrent lors des entrevues est « Ne pas se sentir malade du cancer du sein », mais que les traitements, qui sont supposés guérir les femmes, les rendent malades. La catégorie « Ne pas se sentir malade du cancer du sein » a émergé en tant que catégorie centrale, car elle était au cœur du processus de transition des perceptions de l'état de santé. Elle rattache les autres catégories entre elles afin d'expliquer ce que les femmes ont vécu. La construction graduelle de la schématisation de la transition, au fur et à mesure de l'analyse des données, a montré que cette catégorie était le chaînon auquel toutes les catégories étaient liées (figure 5.2). C'est cette catégorie qui fait le pont entre les perceptions avant et après l'annonce du diagnostic de cancer du sein. C'est aussi cette catégorie qui lie les autres concepts du processus de la transition comme les étapes du deuil, le changement de conception de la vie et les réactions aux représentations du cancer (Strauss et Corbin, 1998).

2.4.4. Présentation et interprétation des résultats

Les données révèlent que les participantes traversent les mêmes étapes que le deuil (choc, déni, colère et marchandage, tristesse, résignation, acceptation et reconstruction) lorsqu'elles reçoivent un diagnostic de cancer du sein. Par la suite, elles vont développer une nouvelle conception de la vie en vivant le moment présent, en étant positives et en changeant leurs priorités. Cela s'illustre comme suit: « *J'étais une personne qui disait: "Je vais économiser dans ma vie. Je vais en profiter à ma retraite." Quand tu as un cancer, ta philosophie de vie change. Peut-être que je vais mourir demain matin donc on va en profiter aujourd'hui* » (P-14). De plus, elles ne se sentent pas malades du cancer du sein, car elles ne ressentent pas de symptômes physiques et perçoivent que le cancer peut être guéri par la chirurgie, comme le mentionne la participante suivante: « *Dans ma tête, il est enlevé. Il n'y en a plus* » (P-11). À l'inverse, elles disent se sentir malades lorsqu'elles ressentent des effets secondaires de la chimiothérapie et de la radiothérapie tels que les changements sur les plans corporel et sexuel comme le gain de poids, les cicatrices, la diminution du sentiment de féminité. Toutefois, elles ne les associent pas à leur perception de la maladie.

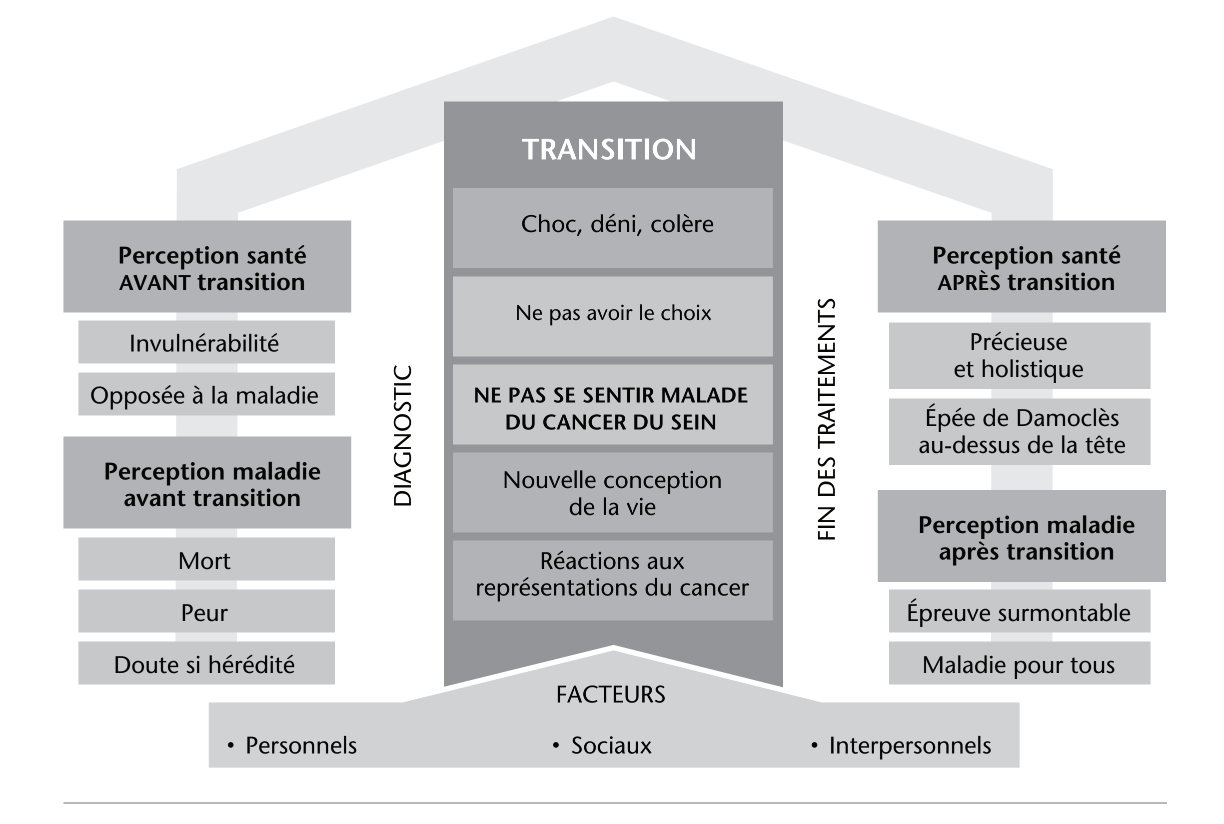
Les participantes vont également changer leurs conceptions de la vie telle qu'avoir une attitude positive. Elles utilisent également la comparaison pour s'encourager dans cette épreuve en se répétant des maximes telles que «Ça aurait pu être pire» et «Quand on se regarde, on se désole; quand on se compare, on se console». Nous avons également remarqué que les participantes qui ont, par le passé, vécu un épisode de dépression vont comparer le cancer aux maladies mentales. En effet, elles affirment que la maladie mentale se vivait plus durement que la maladie physique, car elles se sentaient plus stigmatisées et qu'elles n'avaient pas l'énergie d'être positives (Shim et Rust, 2013). Cet aspect comparatif est important puisque, selon elles, la guérison du corps part de l'esprit, comme le mentionne cette participante: «*Que ce soit un cancer ou autre chose, je pense que le secret, c'est d'être positive*» (P-19). Voici comment l'explique une participante: «*La dépression, je trouve ça pire que le cancer du sein parce que ça affecte toutes les sphères de ta vie parce que tu ne vois plus avec les mêmes lunettes*» (P-10).

Dans le processus de transition des perceptions de l'état de santé, le fait que les participantes ne se sentent pas malades du cancer du sein influe également sur leurs réactions aux représentations du cancer des autres. La participante suivante explique bien cette idée:

> *Tu as l'étiquette de la cancéreuse. Veut veut pas, je n'ai plus de cheveux, j'ai les yeux cernés à cause de la chimio donc les personnes t'approchent comme étant une personne cancéreuse, une personne en traitement, une personne souffrante, une personne fragile. Tu sais la personne malade typique qu'on voit à la télévision. Les gens nous approchent comme ça* (P-28).

Le processus de transition des perceptions de l'état de santé des participantes est illustré à la figure 5.2. Les perceptions de l'état de santé avant et après la transition comme telle sont présentées. La transition proprement dite est composée d'une catégorie centrale, soit le fait de ne pas se sentir malade du cancer du sein. Le diagnostic et la fin des traitements délimitent la transition. Les facteurs qui influent sur le processus de transition complètent la modélisation du processus. Les résultats détaillés font l'objet d'une publication scientifique (Hébert, Gallagher et St-Cyr Tribble, soumis).

Figure 5.2.
Processus de transition des perceptions de l'état de santé de femmes diagnostiquées d'un cancer du sein



Ces quelques exemples de résultats découlant de la théorisation ancrée éclairent les professionnels de la santé sur le processus de transition des perceptions de l'état de santé des participantes. Ils permettent également de mieux comprendre les phases transitoires afin d'adapter les soins et de donner de l'information.

La théorisation ancrée a guidé les chercheuses dans le développement d'une théorie sur un processus de transition. Cette démarche inductive a favorisé l'établissement d'une relation de confiance qui s'est répercutée sur la profondeur des entrevues et, ainsi, sur la richesse des données amassées. La flexibilité de la théorisation ancrée a également permis de moduler les questions de recherche en fonction de l'échantillonnage théorique pour mieux comprendre le phénomène étudié et proposer une théorisation.

CONCLUSION

Dans le cadre de ce chapitre, les principes fondamentaux de la théorisation ancrée ont été présentés et illustrés, tels que la circularité, l'échantillonnage théorique, la comparaison constante et la contextualisation. Nous avons fait référence aux ouvrages fondateurs de la théorisation ancrée, mais il existe d'autres méthodologues qui offrent des variantes à l'approche présentée ici, dont Adèle Clarke (2003) et Pierre Paillé (1994). La théorisation ancrée est, en fait, une occasion d'offrir une nouvelle explication de paradoxes sociaux et de proposer une contribution théorique fortement enracinée dans les données.

RÉFÉRENCES

BIRKS, M. et J. MILLS (2011). *Grounded Theory – A Practical Guide*, Thousand Oaks, Sage Publications.

BLUMER, H. (1969). *Symbolic Interactionism : Perspective and Method*, Englewood Cliffs, Prentice Hall.

BRINGER, J.D., L.H. JOHNSTON et C.H. BRACKENRIDGE (2004). « Maximizing transparency in a doctoral thesis : The complexities of writing about the use of QSR*NVIVO within a grounded theory study », *Qualitative Research*, vol. 4, n° 2, p. 247-265.

BRYANT, A. et K. CHARMAZ (2007). *The Sage Handbook of Grounded Theory*, Thousand Oaks, Sage Publications.

CHARMAZ, K. (2000). « Grounded theory : Objectivist and constructivist methods », dans N.K. Denzin et Y.S. Lincoln (dir.), *Handbook of Qualitative Research*, 2e éd., Thousand Oaks, Sage Publications, p. 509-535

CHARMAZ, K. (2005). « Grounded theory in the 21th century », dans N. K. Denzin et Y. Lincoln (dir.), *The Sage Handbook of Qualitative Research*, 3e éd., Thousand Oaks, Sage Publications, p. 507-535.

CHARMAZ, K. (2006). *Constructing Grounded Theory. A Practical Guide through Qualitative Analysis*, Thousand Oaks, Sage Publications.

CHARMAZ, K. (2011). « Grounded theory methods in social justice research », dans N.K. Denzin et Y.S. Lincoln (dir.), *The Sage Handbook of Qualitative Research*, 4e éd., Thousand Oaks, Sage Publications, p. 359-380.

CLARKE, A.E. (2003). « Situational analyses : Grounded theory mapping after the postmodern turn », *Symbolic Interaction*, vol. 26, p. 553-576.

COONEY, A. (2010). « Choosing between Glaser and Strauss : An example », *Nurse Researcher*, vol. 17, n° 4, p. 18-28.

CORBIN, J. et A. STRAUSS (2008). *Basics of Qualitative Research*, 3e éd., Thousand Oaks, Sage Publications.

CROOKS, V.A., V. CHOUINARD et R.D. WILTON (2008). « Understanding, embracing, rejecting : Women's negotiations of disability constructions and categorizations after becoming chronically ill », *Social Science et Medicine*, vol. 67, p. 1837-1846.

GHEZELJEH, T. N. et A. EMAMI (2009). « Grounded theory : Methodology and philosophical perspective », *Nurse Researcher*, vol. 17, nº 1, p. 15-23.

GLASER, B.G. (1992). *Basics of Grounded Theory Analysis*, Mill Valley, Cal., The Sociology Press.

GLASER, B.G. (2007). « Doing formal theory », dans A. Bryant et K. Charmaz (dir.), *The Sage Handbook of Grounded Theory*, Thousand Oaks, Sage Publications, p. 97-113.

GLASER, B.G. et A.L. STRAUSS (1967). *The Discovery of Grounded Theory : Strategies for Qualitative Research*, New York, Aldine.

HALL, H., D. GRIFFITHS et L. MCKENNA (2013). From Darwin to constructivism : The evolution of grounded theory. *Nurse Researcher*, vol. 20, nº 3, p. 17-21.

HEATH, H. et S. COWLEY (2004). « Developing grounded theory approach : A comparison of Glaser and Strauss », *International Journal of Nursing Studies*, vol. 41, nº 2, p. 141-150.

HÉBERT, M. et M. CÔTÉ (2011). « Étude descriptive rétrospective des perceptions de la maladie de femmes nouvellement atteintes d'un cancer du sein à la suite de leurs traitements en clinique ambulatoire d'oncologie », *L'infirmière clinicienne*, vol. 8, nº 1, p. 1-8.

HÉBERT, M., F. GALLAGHER et D. ST-CYR TRIBBLE (soumis). « La théorisation enracinée dans l'étude de la transition des perceptions de l'état de santé de femmes atteintes d'un cancer du sein ».

HOOD, J.C. (2007). « Orthodoxy vs. Power : The defining traits of grounded theory », dans A. Bryant et K. Charmaz (dir.), *The Sage Handbook of Grounded Theory*, Thousand Oaks, Sage Publications, p. 302-328.

HUNTER, A., K. MURPHY, A. GREALISH, D. CASEY et J. KEADY (2011). « Navigating the grounded theory terrain, part 1 », *Nurse Researcher*, vol. 18, nº 4, p. 6-10.

HUTCHISON, A.J., L.H. JOHNSTON et J.D. BRECKON (2010). « Using QSR-NVivo to facilitate the development of a grounded theory project : An account of a worked example », *International Journal of Social Research Methodology*, vol. 13, nº 4, p. 283-302.

LABELLE, F., O. NAVARRO-FLORES et J. PASQUERO (2013). « Choisir et tirer parti de la méthodologie de la théorisation enracinée », dans J. Luckerhoff et F. Guillemette (dir.), *Méthodologie de la théorisation enracinée : Fondements, procédures et usages*, Québec, Presses de l'Université du Québec.

LAPERRIÈRE, A. (1997). « La théorisation ancrée (*grounded theory*) : démarche analytique et comparaison avec d'autres approches apparentées », dans J. Poupart, J.-P. Deslauriers, L.-H. Groulx, A. Laperrière, R. Mayer et A.P. Pires (dir.), *La recherche qualitative : enjeux épistémologiques et méthodologiques*, Boucherville, Gaëtan Morin, p. 309-340.

LEVENTHAL, H., Y. BENYAMINI, S. BROWNLEE, M. DIEFENBACH, E. A. LEVENTHAL, L. PATRICK-MILLER *et al*. (1997). « Illness representations : Theoretical foundations », dans K. J. Petri et J. A. Weinman (dir.), *Perceptions of Health and Illness. Current Research and Applications*, Canada : Harwood Academic Publishers.

LUCKERHOFF, J. et F. GUILLEMETTE (dir.) (2013). *Méthodologie de la théorisation enracinée*, Québec, Presses de l'Université du Québec.

MASSÉ, R. (1995). *Culture et santé publique*, Boucherville, Gaëtan Morin.

MELEIS, A.I. (2010). *Transitions Theory*, New York, Springer.

MELEIS, A.I., L.M. SAWYER, E. IM, D.K.H. MESSIAS et K. SCHUMACHER (2000). « Experiencing transitions : An emerging middle-range theory », *Advances in Nursing Science*, vol. 23, nº 1, p. 12-28.

MORSE, J. (2007). « Sampling in grounded theory », dans A. Bryant et K. Charmaz (dir.), *The Sage Handbook of Grounded Theory*, Thousand Oaks, Sage Publications, p. 229-244.

MORSE, J.M. (1994). « Emerging from the data : The cognitive process of analysis in qualitative enquiry », dans J.M. Morse (dir.), *Critical Issues in Qualitative Research Methods*, Thousand Oaks, Sage Publications, p. 220-235.

ORGANISATION MONDIALE DE LA SANTÉ – OMS (2013). *Octobre, un mois pour sensibiliser au cancer du sein*, <http://www.who.int/cancer/events/breast_cancer_month/fr/>, consulté le 4 octobre 2013.

PAILLÉ, P. (1994). « L'analyse par théorisation ancrée », *Cahiers de recherche sociologique*, vol. 23, p. 147-181.

PARK, K., S. J. CHANG, H. C. KIM, E. C. PARK, E. S. LEE et C. M. NAM (2009). « Big gap between risk perception for breast cancer and risk factors : Nationwide survey in Korea », *Patient Education and Counseling*, vol. 76, nº 1, p. 113-119.

PIERRON, J.-P. (2007). « Représentations du corps malade et symbolique du mal : maladie, malheur, mal ? », *Psycho-Oncology*, vol. 1, p. 31-40.

PIRES, A.P. (1997). « Échantillonnage et recherche qualitative : essai théorique et méthodologique », dans J. Poupart, J.-P. Deslauriers, L.-H. Groulx, A. Laperrière, R. Mayer et A.P. Pires (dir.), *La recherche qualitative : enjeux épistémologiques et méthodologiques*, Boucherville, Gaëtan Morin, p. 113-169.

RICHARDS, L. et J.M. MORSE (2007). *Read Me First for a User's Guide to Qualitative Methods*, 2e éd., Thousand Oaks, Sage Publications.

SHIM, R. et G. RUST (2013). « Primary care, behavioral health, and public health : Partners in reducing mental health stigma », *American Journal of Public Health*, vol. 103, nº 5, p. e1-e3.

STATISTIQUE CANADA (2011). *Le cancer au Canada : cancers du poumon, du côlon et du rectum, du sein et de la prostate*, <http://www.statcan.gc.ca/pub/82-624-x/2011001/article/11596-fra.pdf>, consulté le 12 juin 2013.

STRAUSS, A. et J. CORBIN (1998). *Basics of Qualitative Research*, 2e éd., Thousand Oaks, Sage Publications.

STRAUSS, A.L. et J.M. CORBIN (1990). *Introduction to Qualitative Research Grounded Theory Procedures and Techniques*, Californie, Sage Publications.

TAN, J. (2010). « Grounded theory in practice : Issues and discussion for new qualitative researchers », *Journal of Documentation*, vol. 66, nº 1, p. 93-112.

WUEST, J. (2012). « Grounded theory : The method », dans P.L. Munhall (dir.), *Nursing Research : A Qualitative Perspective*, 5e éd., Toronto, Jones and Barlett Learning, p. 225-256

CHAPITRE 6

L'ANALYSE DE CONCEPT
Description et illustration de la charge de travail mentale

Valérie Tremblay-Boudreault
Marie-José Durand
Marc Corbière

FORCES

- Elle permet de clarifier un concept ambigu ou flou.
- Elle contribue au développement d'une pensée critique et d'une rigueur intellectuelle.
- Elle permet d'adopter un processus rigoureux dans la conception ou le choix d'un instrument de mesure.

LIMITES

- Elle mène à un produit qui n'est pas final en soi en raison de l'aspect évolutif des concepts.
- Elle dépend des compétences linguistiques de ceux qui entreprennent la démarche.
- Elle est sujette aux biais liés aux idées préconçues des auteurs qui la réalisent.

Les concepts occupent une place prépondérante chez les étudiants, les cliniciens et les chercheurs, notamment dans le champ de la réadaptation en santé mentale. Certains concepts plus abstraits véhiculés dans le domaine de la santé mentale, tels que la qualité de vie, le bien-être ou encore la stigmatisation, ne font pas toujours consensus. Cette absence ou manque de consensus rend difficile la sélection des instruments de mesure pour une évaluation clinique ou dans le cadre d'un projet de recherche. Or certaines méthodes permettent de mieux saisir un concept difficile à opérationnaliser ; l'analyse de concept est l'une d'entre elles.

La première partie de ce chapitre sera consacrée à la présentation de l'analyse de concept, plus particulièrement à la méthode proposée par Walker et Avant (2005). La seconde partie consistera en une illustration, étape par étape, de celle-ci, afin de guider le lecteur vers la réalisation concrète d'une analyse de concept. Plus précisément, cette illustration portera sur la conception d'un outil de mesure sur la charge de travail mentale, et ce, dans le cadre du processus de retour au travail d'un travailleur ayant vécu un trouble mental transitoire.

1. PRÉSENTATION DE LA MÉTHODE

1.1. Brèves définitions et applications de l'analyse de concept

La notion de concept renvoie à une représentation mentale qui permet de comprendre un phénomène et de cerner sa présence dans la réalité (Bell *et al.*, 1996). L'analyse de concept a été décrite par certains auteurs comme étant une méthode rigoureuse et pragmatique pour définir un concept et ainsi contribuer à l'avancement des connaissances (McDonald, 2011). Elle s'est avérée utile à plusieurs égards (Hupcey *et al.*, 1996 ; McKenna et Cutcliffe, 2005), notamment pour 1) clarifier un concept surutilisé ou préciser un concept ambigu ; 2) comprendre plus en profondeur un concept issu d'une théorie ; 3) faciliter la sélection d'un instrument de mesure ; 4) contribuer à la conception d'un instrument de mesure ou d'un guide d'entrevue ; et 5) faire une synthèse de l'état des connaissances.

1.2. Différentes approches de l'analyse de concept

Le pionnier de l'analyse de concept serait J. Wilson (1963). D'abord pensée pour le domaine de l'éducation, sa méthode propose un ensemble de 11 techniques considérées comme étant des outils pédagogiques, lesquelles

peuvent contribuer au développement de la pensée critique des étudiants (Wilson, 1963). Cette méthode sera finalement adaptée au domaine de la santé et elle deviendra particulièrement prépondérante au sein du développement des théories en sciences infirmières (McKenna et Cutcliffe, 2005). Son utilisation s'étend également à d'autres contextes, comme la réadaptation en gériatrie et en santé mentale (Larivière, 2008 ; Levasseur, Tribble et Desrosiers, 2006 ; Tremblay-Boudreault, 2012).

Le tableau 6.1 présente les principales méthodes de l'analyse de concept en sciences de la santé, ainsi que des exemples d'études les ayant utilisées.

Tableau 6.1.
Principales approches d'analyse de concept

Auteurs de la méthode	Exemples d'études
Chinn et Kramer (2008)	Souffrance (Liu, Hsieh et Chin, 2007)
Morse (1995)	Qualité de vie (Plummer et Molzahn, 2009)
Rodgers (2000)	Travail émotionnel (Truc, Alderson et Thompson, 2009)
Schwartz-Barcott et Kim (2000)	Conscience de la situation (Sitterding *et al.*, 2012)
Walker et Avant (2005)	Perception (McDonald, 2011)
Wilson (1963)	Attachement parents-enfant (Bell *et al.*, 1996)

Chacune des méthodes présentées au tableau 6.1 comprend un ensemble d'étapes qui ont certaines similitudes, mais aussi certaines variations dans leur séquence et leur nature. L'ensemble de ces éléments est appuyé par les fondements philosophiques préconisés par les auteurs. Les étapes similaires incluent généralement la justification du choix du concept, la définition du but de l'analyse, l'établissement des différents usages du concept et de ses attributs, l'utilisation de cas, ainsi que le recours à une recension des écrits. La méthode de Walker et Avant (2005) a été retenue pour ce chapitre, car, d'une part, elle est l'une des plus répandues et, d'autre part, parce que les étapes qui la constituent sont claires et accessibles pour les novices (Fulton, Miller et Otte, 2012). Le lecteur intéressé à comprendre et à comparer plus en profondeur les différentes méthodes (fondements philosophiques, étapes, principes des procédures et résultats à l'issue des méthodes) est invité à consulter les travaux sur le sujet (Endacott, 1997 ; Fulton *et al.*, 2012 ; Hupcey *et al.*, 1996 ; McKenna et Cutcliffe, 2005).

1.3. Description de la méthode de Walker et Avant

La méthode de Walker et Avant (2005) comprend huit étapes systématiques présentées dans le tableau 6.2. Chacune d'elles sera reprise de façon détaillée dans les prochaines pages. Ces étapes se déroulent de façon itérative pour ainsi assurer une plus grande précision des résultats de l'analyse, et ce, au fur et à mesure que de nouvelles informations sont obtenues (Walker et Avant, 2005).

Tableau 6.2.
Étapes de l'analyse de concept selon Walker et Avant (2005)

1. Choisir le concept.
2. Définir le but de l'analyse.
3. Établir les différents usages.
4. Déterminer les attributs du concept.
5. Identifier un cas modèle.
6. Identifier un cas contraire, relié, limite, inventé et illégitime.
7. Identifier les antécédents et les conséquents du concept.
8. Identifier les référents empiriques.

1.3.1. Choisir le concept

Le choix du concept est la première étape de l'analyse de concept et elle doit être faite avec soin. Walker et Avant (2005) suggèrent de choisir un concept dans son champ d'intérêt, afin que les résultats qui découlent de l'analyse contribuent significativement à l'avancement des connaissances.

1.3.2. Définir le but de l'analyse

La seconde étape répond à la question « Pourquoi désire-t-on réaliser une analyse de concept ? ». Cette étape aide à demeurer centré sur le but de l'analyse afin que les résultats obtenus soient utiles. Walker et Avant (2005) précisent qu'il est important d'avoir toujours en tête le but de l'analyse, car celui-ci influencera certaines décisions au cours du processus d'analyse. Par exemple, il pourrait être possible de ne retenir qu'un certain nombre d'usages d'un concept (établis à l'étape 3) pour que les résultats soient plus utiles pour le chercheur. Cela pourrait être le cas du concept de « stress », qui revêt parfois une notion de stresseur ou celle de réaction à un stresseur. Si notre étude s'intéresse plus précisément à définir la réaction de stress, seul cet usage pourrait être considéré dans les étapes subséquentes.

1.3.3. Établir les différents usages du concept

C'est à cette étape que les différents usages du concept sont recensés dans les écrits. Les usages représentent les utilisations ou les différentes définitions données à un concept. Il est conseillé d'utiliser des dictionnaires, des livres de référence, des banques de données, des sites Internet, et de consulter les collègues, afin de relever le plus d'usages possibles du concept étudié (Walker et Avant, 2005). À ce stade de l'analyse, il est préférable de ne pas se limiter à un aspect du concept, mais plutôt de retenir tous les usages repérés dans les écrits dans différents domaines et disciplines, afin d'établir un bilan exhaustif. En complément à cette étape, le lecteur est invité à consulter le chapitre portant sur la revue de littérature systématique (voir le chapitre 7 de cet ouvrage). En effet, la revue de littérature systématique présente certaines similarités avec l'établissement des usages du concept, notamment l'utilisation de mots clés pour cibler les écrits pertinents.

1.3.4. Déterminer les attributs du concept

Pour Walker et Avant (2005), déterminer les attributs s'effectue principalement à partir des informations extraites lors de la lecture des écrits recensés et des définitions obtenues dans les dictionnaires. Les attributs correspondent aux caractéristiques qui reviennent le plus souvent pour décrire le concept et ainsi en permettre une définition fine. Concrètement, une liste d'attributs est constituée au fur et à mesure que les informations sont extraites des écrits. Il est important d'aller au-delà des définitions du dictionnaire, sinon les résultats de l'analyse seront superficiels et peu utiles au domaine d'intérêt (Hupcey *et al.*, 1996).

1.3.5. Identifier un cas modèle

Un cas modèle est un exemple qui illustre l'ensemble des attributs du concept (Walker et Avant, 2005). Le cas modèle peut être tiré d'un exemple réel comme de l'expérience clinique, des écrits spécialisés dans le domaine d'intérêt ou même inventé. Bien que cette étape se déroule à la suite de la définition des attributs dans la présentation de la démarche de Walker et Avant (2005), les auteurs proposent diverses alternatives : 1) commencer l'analyse à partir d'un cas modèle ; 2) développer ce cas en même temps que les attributs ; ou 3) faire émerger ce cas après la définition des attributs.

1.3.6. Identifier d'autres types de cas

Examiner d'autres types de cas qui sont similaires ou différents du concept étudié peut aider à préciser ses attributs (Walker et Avant, 2005). Les autres types de cas incluent les cas limite, relié, contraire, inventé et illégitime. Encore une fois, ces cas peuvent provenir d'une situation réelle, des écrits ou être inventés.

Plus précisément, le *cas limite* est défini comme un exemple qui présente certains attributs, voire tous, mais qui diffère de façon évidente pour l'un d'entre eux en intensité ou en durée (Walker et Avant, 2005).

Pour sa part, le *cas relié* correspond à un exemple de concept qui diffère de celui étudié en ne présentant que certains de ses attributs (Walker et Avant, 2005). Le cas relié aide à définir les relations entre le concept étudié et d'autres concepts au sein d'une théorie.

Le *cas contraire* est littéralement un exemple qui ne correspond pas aux attributs du concept analysé. Il peut sembler anodin, voire futile de présenter un exemple de cas contraire. Cet exercice est plutôt profitable pour ce que Wilson (1963) appelle le dialogue interne, soit le processus de réflexion qui sert à confronter les idées et, par conséquent, à confirmer les attributs conjointement avec les autres types de cas.

Le *cas inventé*, quant à lui, est souvent appliqué à la science-fiction afin d'illustrer une application du concept dans un contexte qui sort de l'expérience des auteurs de l'analyse de concept. Ce type de cas est utile lorsque l'analyse porte sur un concept aussi commun que « air » ou « amour » (Walker et Avant, 2005).

Finalement, le *cas illégitime* est un exemple d'usage du concept qui est inadéquat ou qui n'est pas approprié au contexte d'intérêt (Walker et Avant, 2005). Ce type de cas est utile lorsqu'un usage du concept diffère complètement des autres usages recensés dans les écrits. Certains des attributs définis pourraient ne pas s'appliquer au concept correspondant au cas illégitime.

1.3.7. Identifier les antécédents et les conséquents du concept

Walker et Avant (2005) définissent les antécédents comme les événements ou les incidents qui se produisent avant l'apparition d'un concept. À l'inverse, les conséquents représentent ce qui résulte de la présence du concept. Ces éléments aident à préciser les attributs du concept, puisqu'un attribut ne peut pas être à la fois un antécédent et un conséquent. De plus, ils permettent de situer le concept d'intérêt dans un univers de concepts

provenant d'un même domaine de connaissances. Ceci permet aussi d'émettre de nouvelles hypothèses de recherche en proposant des variables ou des relations possibles entre des concepts. L'identification des antécédents et des conséquents s'effectue de la même manière que celle des usages et des attributs, c'est-à-dire en extrayant les informations des écrits recensés. Les cadres de référence s'avèrent particulièrement utiles pour identifier les antécédents d'un concept et ses conséquents. Une analyse de concept sur la participation sociale a permis de dégager qu'un être humain en interaction avec un environnement physique et social représente les antécédents de ce concept, alors que le bien-être ou l'adoption de mauvaises habitudes de vie sont des exemples des deux pôles de conséquents de la participation sociale (Larivière, 2008).

1.3.8. Identifier les référents empiriques

Les référents empiriques peuvent être identifiés dans les écrits en se posant des questions telles que « Comment les auteurs mesurent-ils le concept ? » ou « Comment la présence d'un concept se manifeste-t-elle dans la vraie vie ? ». En d'autres mots, les référents empiriques sont des indicateurs de la présence du concept et, par conséquent, de celle de ses attributs. Walker et Avant (2005) précisent que dans certains cas, les attributs peuvent être identiques aux référents empiriques. Les référents empiriques seraient surtout utiles pour illustrer la manifestation de la présence d'un concept très abstrait.

À cette étape, il est utile de refaire une recension des écrits qui se rapportent aux mesures du concept d'intérêt, en ajoutant des mots clés associés en général à la validation d'outils de mesure (p. ex., validation, validité, test-retest, fidélité, validité factorielle, validité de convergence). En l'occurrence, un bilan critique des outils de mesure pourra être effectué et ainsi orienter le choix vers celui qui est le plus adapté aux besoins du contexte clinique ou de recherche. Une première étape qui s'avère importante pour guider le choix de l'outil de mesure est de comparer sa compatibilité avec les attributs déterminés à l'étape 4 décrite précédemment. Idéalement, l'outil retenu devrait couvrir l'ensemble des attributs. Par la suite, les critères habituels liés aux qualités psychométriques et à la passation d'un test sont également comparés pour la population d'intérêt (p. ex., validité, fidélité, temps de passation). En ce sens, l'identification des référents empiriques peut procurer aux cliniciens une définition claire et opérationnelle d'un concept (Walker et Avant, 2005). Dans l'éventualité où les attributs déterminés ne sont pas entièrement couverts par les référents empiriques actuels, ceux-ci peuvent servir de base pour la conception d'un outil de mesure.

1.4. Matériel nécessaire

L'analyse de concept requiert l'accès à des dictionnaires (linguistiques, médicaux, psychologiques, etc.), à des banques de données et à des revues électroniques afin de réaliser la recension des écrits de façon exhaustive. Un logiciel spécialisé pour les références bibliographiques (p. ex., Refworks, EndNotes) est un aout pour faciliter la gestion d'un plus grand volume d'écrits. Une grille d'extraction conçue pour les besoins de l'étude représente un autre outil pour faciliter l'analyse des données (figure 6.1). Enfin, certaines méthodes d'analyse de concept incluent une composante qualitative dans leur démarche (Morse, 1995 ; Schwartz-Barcott et Kim, 2000), ce qui exige par conséquent un enregistreur audionumérique et idéalement un logiciel d'analyse de contenu (p. ex., NVivo) en vue de traiter les données recueillies lors d'entrevues et d'observations participantes.

1.5. Pièges, difficultés et conseils pratiques

Plusieurs pièges et difficultés guettent ceux qui entreprennent une analyse de concept. Le tableau 6.3 présente les principaux écueils qui sont rapportés dans les écrits ou issus de l'expérience de terrain, alors que quelques conseils pratiques sont présentés au tableau 6.4 dans l'objectif de faciliter le processus.

Tableau 6.3.
Pièges, difficultés et écueils de l'analyse de concept

- Avoir l'impression qu'il est facile de réaliser une analyse de concept (Wilson, 1963) ;
- Avoir une tendance compulsive à vouloir tout analyser (Wilson, 1963) ;
- Se sentir désespérément perdu dans le bassin de données (Wilson, 1963) ;
- Présenter des difficultés ou un manque de volonté à débattre avec soi ou avec autrui (Wilson, 1963) ;
- Présenter un cas modèle qui est en contradiction avec les attributs (McKenna et Cutcliffe, 2005) ;
- Vouloir ajouter des attributs si la liste semble trop courte (McKenna et Cutcliffe, 2005) ;
- Désirer confirmer trop rapidement l'idée que l'on a du concept analysé (Rodgers, 2000) ;
- Se limiter aux définitions du dictionnaire (Hupcey *et al.*, 1996) ;
- Présenter un cas sans le discuter, entraînant une incompréhension du lecteur (Hupcey *et al.*, 1996) ;
- Manquer de critères reconnus pour évaluer la qualité scientifique de la méthode (Hupcey *et al.*, 1996) ;
- Confondre l'analyse de concept avec une autre méthode comme la revue systématique ou intégrative des écrits.

Tableau 6.4.
Conseils pratiques pour faire une analyse de concept

- Établir et documenter rigoureusement la démarche de la recension et justifier les critères pour avoir conservé ou rejeté un article (Truc *et al.*, 2009) ;
- Établir une stratégie d'échantillonnage pour documenter le processus de sélection et limiter le nombre des écrits (Rodgers, 2000 ; Truc *et al.*, 2009) ;
- Utiliser une grille d'extraction des données et préparer des questions pour aider à classer les informations dans chacune des cases de la grille (Rodgers, 2000 ; Truc *et al.*, 2009) ;
- Prétester la grille d'extraction et poursuivre la coextraction jusqu'à ce que le pourcentage d'accord ait été jugé satisfaisant par l'équipe selon le seuil fixé au départ ;
- Documenter toute la démarche de l'analyse de concept, incluant les décisions prises, les réflexions et les idées préconçues quant au concept (Morse, 1995 ; Schwartz-Barcott et Kim, 2000) ;
- Discuter régulièrement des résultats avec des collègues (les initiés comme les non-initiés au phénomène d'intérêt) après chaque phase de l'analyse ou au besoin ;
- Présenter clairement les limites de l'analyse de concept et en quoi elle se différencie des autres méthodes (p. ex., recension systématique ou intégrative des écrits) au moment de soumettre l'article à l'éditeur de la revue afin d'éviter la confusion des pairs lors du processus de révision ;
- Cibler des revues qui ont déjà publié ce type d'analyse pour augmenter les chances de publication.

Selon Wilson (1963), ces pièges et difficultés font partie du processus d'apprentissage d'une nouvelle méthode. Ainsi, il est normal de vivre certaines difficultés, voire l'ensemble de ces difficultés au cours du processus. Pour le moins, le lecteur sera avisé des éventuels écueils qu'il pourra rencontrer lors de l'utilisation de l'analyse de concept. Il sera également à même d'appliquer certaines des stratégies proposées dans le tableau 6.4 pour réaliser la démarche.

Avant d'illustrer la méthode de Walker et Avant (2005), un conseil pratique nécessite d'être approfondi. Rodgers (2000) suggère d'utiliser des questions en s'appropriant le contenu des articles pour favoriser une extraction neutre des informations, telles que « Qu'est-ce que l'auteur considère comme un antécédent du concept ? » ou « Comment l'auteur définit-il le concept ? » ou encore « Quels mots l'auteur utilise-t-il pour décrire le concept ? ». En d'autres termes, se poser ces questions au moment de la lecture des articles permet de mettre de côté ses idées préconçues. Rodgers (2000) propose également d'utiliser une grille d'extraction pour chaque catégorie d'informations (p. ex., attributs, antécédents) dans le but d'organiser les informations recueillies. Une autre possibilité est d'utiliser, pour chaque article, une nouvelle grille qui intègre les questionnements propres à chaque catégorie d'information recherchée (p. ex., attributs, conséquents, référents empiriques). Un exemple de grille par article est présenté à la figure 6.1. Pour assurer la rigueur et la qualité de l'extraction des informations, celle-ci peut être effectuée par deux personnes jusqu'à ce que les résultats soient jugés suffisamment concordants.

Figure 6.1.
Exemple de grille d'extraction

N° de l'article	*Ces cases sont utiles pour documenter la référence de l'article et les principaux résultats.*
Référence	
But de l'article	
Domaine	
Résultats (si pertinents)	
Usages	*Dans quels mots les auteurs définissent-ils le concept ?*
Attributs	*Quelles sont les caractéristiques employées par les auteurs pour décrire le concept ?*
Antécédents	*Selon les auteurs, quels sont les éléments qui précèdent la présence du concept ?*
Conséquences	*Quelles sont les conséquences de la présence du concept ?*
Référents empiriques (indicateurs)	*Comment les auteurs mesurent-ils la présence du concept ?*
Référents empiriques (critères)	*Quels sont les critères établis pour mesurer adéquatement le concept ?*
Concepts apparentés et leurs définitions	*Quels sont les concepts apparentés et comment sont-ils définis par les auteurs ?*
Exemples de cas (p. ex. modèle ou autres)	*Les auteurs proposent-ils un exemple de cas illustrant le concept ? Si oui, quel est-il ?*
Commentaires des auteurs	*Ces cases servent à noter les commentaires des auteurs des articles ou de ceux qui réalisent l'analyse de concept.*
Commentaires personnels/réflexion	

Source : Tremblay-Boudreault, 2012.

2. ILLUSTRATION DE LA MÉTHODE : ANALYSE DE CONCEPT DE LA CHARGE DE TRAVAIL MENTALE

Cette partie sera consacrée à l'illustration pratique de l'analyse de concept de la charge de travail mentale dans le contexte de personnes qui sont en processus de retour au travail à la suite d'un trouble mental transitoire (Tremblay-Boudreault, 2012). Les étapes de l'analyse de concept proposée par Walker et Avant (2005) seront utilisées pour soutenir cette illustration.

2.1. Choisir le concept

Le concept de charge de travail mentale (CTM) a été retenu, car un niveau inadéquat de CTM représente un obstacle au retour au travail à la suite d'un trouble mental transitoire (TMT) (Durand *et al.*, 2010 ; van Oostrom *et al.*, 2007). Les TMT font référence à une classification des troubles mentaux acceptée en réadaptation au travail et regroupent généralement trois catégories : 1) troubles de l'adaptation ; 2) troubles de l'humeur (p. ex., dépression majeure) ; et 3) troubles anxieux (p. ex., trouble de l'anxiété généralisée) (van der Klink et van Dijk, 2003). La mesure de la CTM pourrait permettre une meilleure planification des tâches lors du retour au travail progressif à la suite d'un TMT. Cependant, l'absence de consensus quant à la définition de la CTM et les lacunes reliées à sa mesure ne permettent pas aux cliniciens en réadaptation au travail en santé mentale de l'évaluer de manière rigoureuse. Une définition opérationnelle et rigoureuse de la CTM relative à ce contexte particulier représenterait une valeur ajoutée aux connaissances scientifiques et cliniques (Walker et Avant, 2005).

2.2. Définir les buts de l'analyse

Le but de la présente analyse est d'abord de bien définir la CTM dans le processus de retour au travail à la suite d'un TMT. Il est à noter que cette démarche s'inscrivait dans un projet plus vaste visant une définition de la CTM sous la forme opérationnelle d'un questionnaire.

2.3. Établir les différents usages

La troisième étape vise à établir tous les usages de la CTM. À titre d'exemple, si l'on fait une recherche avec « *mental workload* » comme mot clé dans la banque de données *Google Scholar*, on obtient 17 500 résultats en date du 6 juillet 2012 (sans restriction de temps) et 667 résultats seulement pour l'année 2012. Si on ajoute le mot clé « *return-to-work* », les résultats obtenus,

pour la même date, sont respectivement 150 et 8. Le choix des mots clés et la période couverte par la recension sont donc des critères essentiels pour bien circonscrire la démarche (voir le chapitre 7 sur la revue systématique des écrits).

En ce qui concerne la CTM, la recension des écrits a été effectuée en utilisant des mots clés précis (p. ex., « charge de travail mentale », « charge cognitive », « charge mentale », « charge émotionnelle », « charge affective », « évaluation », « mesure » et leurs équivalents en anglais) dans des dictionnaires, des banques de données (Medline, Embase, PsycINFO, Eric, CINAHL, ProQuest Dissertations and Thesis), ainsi que dans des ouvrages de référence et des sites Internet sur la santé au travail. La variété des sources de données a permis en premier lieu d'obtenir un portrait plus global des différents usages du concept de la CTM, tout en retenant par la suite les caractéristiques les plus pertinentes au contexte ciblé. La liste des références bibliographiques des documents recensés a également été consultée afin de repérer des sources d'informations complémentaires. La recherche a été limitée aux articles en anglais et en français publiés dans la période de 2000 à 2010. Ces choix reposent, d'une part, sur la volonté de vouloir limiter les erreurs de compréhension dues à la barrière de la langue et, d'autre part, sur une période de temps récente, mais assez large, tout en ayant la possibilité de consulter les références bibliographiques des articles retenus. Afin d'extraire les informations pertinentes des articles retenus, une grille basée sur les étapes de l'analyse de concept (définitions proposées, attributs, cas modèle et autres types de cas, antécédents, conséquents et référents empiriques) a été conçue pour chaque article en tentant de répondre à des questions précises (p. ex., « Qu'est-ce que l'auteur entend par CTM ? »). Un extrait de la grille remplie pour les différents usages est présenté à la figure 6.2.

Au terme de cette étape, 27 définitions de la CTM ont été répertoriées, en plus des définitions de chacun des mots du concept recherchés séparément, puis ensemble, dans les différents dictionnaires. Malgré le nombre important de définitions disponibles, deux principaux usages de la CTM semblaient se dégager des écrits (Organisation internationale de normalisation, 1991). Le premier usage est inhérent au travail (condition du travail) lorsque la CTM est considérée sous l'aspect de la contrainte, notamment la charge anticipée lors de la conception d'un poste. Un exemple de définition liée à la condition de travail correspond à « la somme des exigences placées sur l'opérateur » (Bridger, 2009, p. 480 ; traduction libre). Le second usage de la CTM concerne la réalisation d'une activité de travail lorsque la CTM est considérée sous l'aspect de l'astreinte, incluant la charge telle que vécue par l'individu. La définition qui illustre cette seconde perspective va comme suit : « l'astreinte d'un opérateur ou l'effort associé à l'accomplissement d'une

tâche» (Megaw, 2005, p. 526; traduction libre). Il y a donc une nuance entre les définitions, celles centrées sur la tâche et celles centrées sur l'individu (Hart et Staveland, 1988). Walker et Avant (2005) encouragent de se référer aux buts de l'analyse afin de guider le choix des perspectives à analyser plus en profondeur et ainsi éviter l'obtention de résultats inutiles ou peu pratiques. Le but ultime est de définir opérationnellement la CTM afin que cette définition puisse guider les cliniciens dans le processus de retour au travail de personnes avec un TMT. Considérant que ce processus repose sur une approche personnalisée, seule la perspective de la conséquence de l'activité de travail a été retenue afin de poursuivre les étapes 4 à 7 de l'analyse de concept.

Figure 6.2.
Extrait de la grille d'extraction (usages)

Référence	Leplat, J. (1994). Quelques commentaires sur la notion de charge mentale de travail. *Revue de médecine psychosomatique*, 35(*40*), 27-42.	
…		
Usages	p. 29	«Degré de mobilisation d'un individu pour l'exécution d'une tâche donnée»
	p. 29	«le rapport des exigences de la tâche à la capacité maximale moyenne de l'opérateur ou comme le pourcentage de capacité requis pour répondre à ces exigences (Welford, 1997, p. 284).»
	p. 30	«tout ce qui pèse sur lui, c'est-à-dire ce qui contribue à dégrader son épanouissement, son bien-être, sa santé, son intégrité physique, à court, moyen et long terme…»

Source : Tremblay-Boudreault, 2012.

2.4. Déterminer les attributs

La quatrième étape permet de déterminer les attributs, c'est-à-dire les principales caractéristiques qui désignent un concept. La figure 6.3 présente un extrait de la grille d'extraction reliée à cette étape.

Figure 6.3.
Extrait de la grille d'extraction (attributs)

Attributs	p. 28 « Le travail est une activité et la charge de travail représente ce que coûte cette activité à celui qui la pratique : elle est une conséquence de l'activité pour l'individu ».
	p. 29 « c'est une notion liée à l'interaction de ces deux variables (exigences du travail et capacités) ».
	p. 30-31 Coût cognitif « le coût cognitif désigne l'effort que le sujet doit faire pour acquérir ces nouvelles connaissances et, à partir d'elles, élaborer une autre procédure ».

Source : Tremblay-Boudreault, 2012.

Les informations recueillies à partir de l'ensemble des articles retenus ont par la suite été analysées pour en dégager les attributs. Dans cette illustration qui prend en compte la perspective de la conséquence de l'activité de travail, la CTM consiste en quatre attributs :

- Un état mental immédiat ou à court terme relié à l'effort ou au coût pour l'individu découlant de la réalisation de son activité de travail pour un niveau de performance donné.
- Un état mental qui présente des composantes cognitives et affectives.
- Un continuum, allant de la sous-charge à la surcharge de travail mentale, en passant par une charge confortable.
- Un état mental néfaste pour l'individu en situation de sous-charge ou de surcharge, et un état mental profitable à l'individu en situation de charge confortable.

2.5. Identifier un cas modèle

Le cas modèle se veut une illustration de l'ensemble des attributs du concept. La présentation séquentielle des étapes ayant été suivies, le cas modèle a été développé à partir d'un exemple clinique, et ce, après avoir déterminé les attributs. Le cas modèle va comme suit :

> *Une femme de 32 ans effectue un retour au travail progressif à la suite d'une absence de sept mois liée à un épisode de dépression majeure. Lors des premières journées de travail, elle ne ressent pas de défi personnel pour effectuer les tâches*

qui lui sont confiées. Elle se sent même parfois inutile et elle ne parvient pas à maintenir une attention soutenue. Dès la troisième semaine, elle se sent tendue et sous pression à plusieurs reprises lorsqu'elle réalise ses tâches de travail. Elle ne parvient plus à organiser toutes les informations de son travail de façon efficace. Elle se sent épuisée à la fin de ses journées de travail et elle doute constamment de ses capacités à atteindre les objectifs fixés. Son patron lui confie alors des tâches moins exigeantes pour tenir compte de son retour progressif. Elle n'a plus l'impression que son niveau d'effort est trop élevé. Elle se sent maintenant en contrôle pour effectuer les tâches qui lui sont confiées et elle ressent même du plaisir à faire son travail.

Dans ce cas-ci, le cas modèle débute avec une mise en contexte. Les deux premières semaines du retour au travail de la jeune femme sont caractérisées par un état de sous-charge mentale dont les manifestations sont des expériences cognitives et affectives négatives pour elle. Dès la troisième semaine, elle vit plutôt un état de surcharge de travail mentale. Les expériences cognitives et affectives demeurent néfastes pour elle, même si elles diffèrent de l'état de sous-charge initial. À la suite des modifications apportées aux tâches de son travail, elle fait l'expérience d'un état de charge confortable. En partant d'une situation clinique vécue, la rédaction de ce cas a permis de confirmer le choix des attributs et d'approfondir la réflexion, notamment en créant d'autres types de cas, présentés ci-dessous.

2.6. Identifier d'autres types de cas

2.6.1. Cas limite

Après avoir bu plusieurs verres de vin pendant le repas du midi, un homme est somnolent à son travail et doit faire de grands efforts pour se concentrer sur ses tâches. Il est déçu de lui-même lorsqu'il constate qu'il ne pourra pas respecter les délais prévus pour réaliser son travail.

Dans cet exemple, l'homme expérimente un état mental qui se manifeste sur les plans affectif et cognitif. Ainsi, la plupart des attributs sont présents dans cet exemple, mais le continuum avec les deux extrémités néfastes pour l'individu ne s'y retrouve pas. De plus, l'effet immédiat et à court terme est celui de l'alcool et non de l'effort associé à la tâche elle-même. Cet exemple serait plus approprié pour illustrer le fait d'être sous l'influence de l'alcool. La construction de ce cas a mis en évidence l'importance d'inclure le continuum de la CTM parmi les attributs, plus précisément en rendant explicite le type de conséquence immédiate ou à court terme (néfaste ou profitable) qui caractérise la CTM.

2.6.2. Cas relié

Un homme est convoqué au bureau de son patron dans quinze minutes. Pendant les minutes qui suivent, l'homme ne parvient pas à se concentrer. Il ne fait que craindre le pire. Il a peur que son patron lui annonce qu'il n'est pas suffisamment compétent et d'être congédié sur-le-champ.

Dans cet exemple, bien que l'homme vive des expériences néfastes sur les plans affectif et cognitif, celles-ci ne se présentent pas sur un continuum, mais plutôt comme une anticipation négative de la rencontre avec son patron. Cet exemple illustre plutôt une anxiété vécue dans le cadre du travail, une expérience qui représente des points communs avec la surcharge de travail mentale, mais pour laquelle l'état mental ne provient pas de la réalisation d'une activité de travail. D'où l'importance de bien repérer les antécédents et les conséquents de la CTM (étape suivante). L'utilisation de ce cas a aussi été utile pour confirmer l'importance d'inclure l'attribut précisant que la CTM est un état mental associé à l'effort ou au coût engendré par une activité de travail réalisée selon un niveau de performance donné. Ceci diffère grandement de la notion d'anticipation associée à l'anxiété.

2.6.3. Cas contraire, cas inventé et cas illégitime

Ces trois types de cas sont également proposés dans la démarche de Walker et Avant (2005). Comme la contribution de ces derniers n'était pas pertinente pour l'analyse de la CTM, ces cas n'ont pas été conçus[1]. Par exemple, la CTM n'est pas un concept commun au même titre que «air» ou «amour»; ainsi le cas inventé n'apporterait pas de valeur ajoutée aux résultats de l'analyse. De plus, aucun exemple de cas illégitime (usage inadéquat ou inapproprié au contexte) n'a été soulevé lors de la recension des écrits. Le lecteur peut toujours consulter Walker et Avant (2005) pour des exemples concrets de ces cas.

2.7. Identifier les antécédents et les conséquents

Cette étape permet de se pencher sur le contexte d'utilisation du concept, soit la CTM, en identifiant ses antécédents et ses conséquents. Un extrait de la grille d'extraction correspondant à cette étape est présenté à la figure 6.4.

1. Outre le cas modèle, l'exercice de trouver des exemples de cas demeure à la discrétion du chercheur. Il n'est pas une obligation en soi de trouver des exemples pour chaque type de cas.

Figure 6.4.
Extrait de la grille d'extraction (antécédents et conséquents)

Antécédents	p. 28	« Exigences du travail : contraintes imposées par la tâche et auxquelles l'activité doit répondre. »
	p. 36-37	• conditions internes (caractéristiques de l'individu) – habileté, compétence – automatisme vs compétence – motivation pour l'exécution – traits généraux de personnalité
Conséquents	p. 30	Selon Theureau (1992) : son épanouissement, son bien-être, sa santé, son intégrité physique, à court, moyen et long terme

Source : Tremblay-Boudreault, 2012.

Les antécédents de la CTM qui ont été identifiés sont : 1) la capacité limitée de l'humain à traiter l'information ; 2) les contraintes de l'environnement associées aux exigences de la tâche, aux facteurs de l'environnement physique, aux facteurs organisationnels et aux facteurs sociaux qui affectent mentalement un individu (positivement ou négativement) ; et 3) le comportement chez l'humain qui se traduit par une interaction avec son environnement, incluant ses caractéristiques personnelles telles que la motivation, son niveau de compétence, ses stratégies ou modes opératoires préconisés, son état de santé et d'activation initial, pour n'en nommer que quelques-uns.

Les conséquents à moyen et long terme de la CTM identifiés dans les écrits sont : 1) la sécurité (opérateurs et système) ; 2) la performance, incluant la détérioration et les échecs associés à la surcharge de travail et la sous-charge de travail ; et 3) les effets chroniques sur le bien-être de l'individu (surtout liés à la surcharge de travail mentale).

La réalisation de cette étape a permis de situer les attributs de la CTM par rapport aux autres concepts qui précèdent ou suivent la CTM. Puisqu'un attribut ne devrait pas être également un antécédent ou un conséquent (Walker et Avant, 2005), cette étape permet de préciser une nouvelle fois le choix des attributs.

2.8. Identifier les référents empiriques

La dernière étape de l'analyse de concept est celle où les référents empiriques de la CTM sont identifiés. La figure 6.5 illustre un extrait de la grille d'extraction pour cette étape de l'analyse de concept.

Figure 6.5.
Extrait de la grille d'extraction (référents empiriques)

Référents empiriques (indicateurs)	p. 34	• Indicateurs physiologiques • Indicateurs issus de l'activité de travail • Indicateurs issus du sentiment de charge • Indicateurs issus des procédures indirectes d'évaluation (performance)
Référents empiriques (critères)	p. 34	Selon Eggemeir *et al.* (1991) : • Sensibilité (niveau de charge) • Niveaau d'encombrement (*intrusiveness*) • Exigences de mise en œuvre • Acceptabilité

Source : Tremblay-Boudreault, 2012.

Quatre grandes catégories de référents empiriques ont été recensées pour la CTM (Megaw, 2005 ; Organisation internationale de normalisation, 1991) : 1) les mesures de performance (p. ex., temps pour réaliser la tâche, nombre d'erreurs) ; 2) les mesures physiologiques (p. ex., dilatation de la pupille, variation du rythme cardiaque) ; 3) les mesures issues de l'analyse de l'activité de travail (p. ex., distance entre l'œil et la tâche, variation des modes opératoires) ; et 4) les mesures subjectives comme le questionnaire Nasa Task Load Index, représentant la charge de travail mentale en six dimensions (c.-à-d. exigences physiques, exigences cognitives, exigences émotives, effort, frustration et performance) et le questionnaire Rating Scale Mental Effort, représentant la charge de travail mentale en une seule dimension (c.-à-d. effort).

L'un des constats de cette étape est qu'aucune des mesures existantes de la CTM ne répondait aux critères souhaités pour mesurer ce concept, tel que défini dans une perspective de conséquence du travail. Les mesures subjectives sont celles qui couvrent essentiellement l'ensemble des attributs de la CTM dans le processus de retour au travail à la suite d'un TMT. Toutefois, à notre connaissance, aucune des mesures subjectives n'avait été validée auprès de la population d'intérêt. De plus, celles-ci ne permettaient pas de distinguer les différentes étapes du continuum, plus particulièrement

la sous-charge et la charge confortable, lesquelles sont peu décrites dans les écrits. Par conséquent, il nous semblait important d'amorcer la conception d'un nouveau questionnaire de la CTM adapté au contexte des personnes aux prises avec un trouble mental transitoire en processus de retour au travail. Les attributs ont servi de base pour définir les dimensions du questionnaire. Par la suite, les référents empiriques identifiés dans la catégorie des mesures subjectives ont guidé la rédaction des items pour couvrir chacun des attributs, en tenant compte du continuum de la CTM (sous-charge, charge confortable et surcharge). La contribution de l'analyse de concept au processus de conception d'un questionnaire dépasse les objectifs de ce chapitre et sera développée dans le chapitre 24 de cet ouvrage. Le lecteur pourra aussi consulter la conception du questionnaire de la CTM dans l'étude de Tremblay-Boudreault (2012).

CONCLUSION

L'analyse de concept est une méthode rigoureuse qui permet de clarifier la définition d'un concept. Cette méthode permet notamment de proposer une définition conceptuelle basée sur les attributs du concept étudié. La démarche en huit étapes proposée par Walker et Avant (2005) suggère même d'opérationnaliser le concept au terme de la dernière étape (référents empiriques). Les résultats de l'analyse de concept peuvent également prendre la forme de recommandations quant à la sélection d'un outil de mesure, lorsque des outils existants répondent aux besoins du contexte ciblé. Lorsque ce n'est pas le cas, comme dans l'illustration présentée dans ce chapitre, les résultats de l'analyse de concept fournissent plutôt une base solide pour la conception d'un questionnaire. Cette base représente un fondement théorique au questionnaire et contribue ainsi à la validité de contenu et de construit de celui-ci. Ainsi, entreprendre une analyse de concept demeure une stratégie rigoureuse pour le développement d'outils de mesure ou de théories et cette démarche est pertinente pour l'avancement des connaissances.

RÉFÉRENCES

BELL, L., C. GOULET, D. ST-CYR TRIBBLE, D. PAUL et V. POLOMENO (1996). « Une analyse du concept d'attachement parents-enfant », *Recherche en soins infirmiers*, vol. 46, p. 1-13.

BRIDGER, R.S. (2009). « Human information processing and mental workload », dans R.S. Bridger (dir.), *Introduction to Ergonomics*, 3e éd., Boca Raton, CRC Press Taylor et Francis Group, p. 457-501.

CHINN, P.L. et M.K. KRAMER (2008). *Integrated Theory and Knowledge Development in Nursing*, 7e éd., St-Louis, Elsevier.

DURAND, M.-J., M. CORBIÈRE, C. BRIAND, M.F. COUTU, L. ST-ARNAUD et N. CHARPENTIER (2010). *Les facteurs reliés aux absences prolongées du travail en raison d'un trouble mental transitoire : développement d'un outil de mesure*, no R-674, Montréal, Institut de recherche Robert-Sauvé en santé et en sécurité du travail.

ENDACOTT, R. (1997). « Clarifying the concept of need : A comparison of two approaches to concept analysis », *Journal of Advanced Nursing*, vol. 25, no 3, p. 471-476.

FULTON, J.S., W.R. MILLER et J.L. OTTE (2012). « A systematic review of analyses of the concept of quality of life in nursing : Exploring how form of analysis affects understanding », *Advances in Nursing Science*, vol. 35, no 2, p. E1-12, <doi:10.1097/ANS.0b013e318253728c>.

HART, S.G. et L.E. STAVELAND (1988). « Development of a multi-dimensional workload rating scale : Results of empirical and theoretical research », dans P.A. Hancock et N. Meshkati (dir.), *Human Mental Workload*, Amsterdam, Elsevier Science Publishers B.V., p. 239-250.

HUPCEY, J.E., J.M. MORSE, E.R. LENZ et M.C. TASON (1996). « Wilsonian methods of concept analysis : A critique », *Scholarly Inquiry for Nursing Practice*, vol. 10, no 3, p. 185-210.

LARIVIÈRE, N. (2008). « Analyse du concept de la participation sociale : définitions, cas d'illustration, dimensions de l'activité et indicateurs », *Canadian Journal of Occupational Therapy*, vol. 75, no 2, p. 114-127.

LEVASSEUR, M., D. ST-CYR. TRIBBLE et J. DESROSIERS (2006). « Analysis of quality of life concept in the context of older adults with physical disabilities [Analyse du concept qualité de vie dans le contexte des personnes âgées avec incapacités physiques] », *Canadian Journal of Occupational Therapy*, vol. 73, no 3, p. 163-177.

LIU, G.J., H.Y. HSIEH et C.C. CHIN (2007). « Concept analysis of suffering », *Hu Li Za Zhi the Journal of Nursing*, vol. 54, no 3, p. 92-97.

MCDONALD, S.M. (2011). « Perception : A concept analysis », *International Journal of Nursing Knowledge*, vol. 23, no 1, p. 1-9.

MCKENNA, H.P. et J.R. CUTCLIFFE (2005). « An introduction to concepts and their analyses », dans J.R. Cutcliffe et H.P. McKenna (dir.), *The Essential Concepts of Nursing*, Londres, Elsevier, p. 1-13.

MEGAW, T. (2005). « The definition and measurement of mental workload », dans J.R. Wilson et N. Corlett (dir.), *Evaluation of Human Work*, 3e éd., Boca Raton, CRC Press Taylor et Francis, p. 525-551.

MORSE, J.M. (1995). « Exploring the theoretical basis of nursing using advanced techniques of concept analysis », *Advances in Nursing Science*, vol. 17, no 3, p. 31-46.

ORGANISATION INTERNATIONALE DE NORMALISATION (1991). *ISO 10075. Principes ergonomiques concernant la charge de travail mental : termes généraux et leurs définitions*, Genève, Organisation internationale de normalisation.

PLUMMER, M. et A.E. MOLZAHN (2009). « Quality of life in contemporary nursing theory : A concept analysis », *Nursing Science Quarterly*, vol. 22, nº 2, p. 134-140, <doi :10.1177/0894318409332807>.

RODGERS, B.L. (2000). « Concept analysis : An evolutionary view », dans B.L. Rodgers et K.A. Knafl (dir.), *Concept Development in Nursing*, 2e éd., Philadelphie, W.B. Saunders Company, p. 77-117.

SCHWARTZ-BARCOTT, D. et H.S. KIM (2000). « An expansion and elaboration of the hybrid model of concept development », dans B.L. Rodgers et K.A. Knafl (dir.), *Concept Development in Nursing*, 2e éd., Philadelphie, W.B. Saunders Company, p. 129-159.

SITTERDING, M.C., M.E. BROOME, L.Q. EVERETT et P. EBRIGHT (2012). « Understanding situation awareness in nursing work : A hybrid concept analysis », *Advances in Nursing Science*, vol. 35, nº 1, p. 77-92, <doi :10.1097/ANS.0b013e3182450158>.

TREMBLAY-BOUDREAULT, V. (2012). *Conception d'un outil de mesure de la charge de travail mentale dans le processus de retour au travail à la suite d'un trouble mental transitoire ayant engendré des incapacités prolongées*, Mémoire de maîtrise, Sherbrooke, Faculté de médecine et des sciences de la santé, Université de Sherbrooke.

TRUC, H., M. ALDERSON et M. THOMPSON (2009). « Le travail émotionnel qui sous-tend les soins infirmiers : une analyse évolutionnaire de concept [Emotional labour : An evolutionary concept analysis] », *Recherche en soins infirmiers*, vol. 97, p. 34-49.

VAN DER KLINK, J.J. et F.J. VAN DIJK (2003). « Dutch practice guidelines for managing adjustment disorders in occupational and primary health care », *Scandinavian Journal of Work, Environment et Health*, vol. 29, nº 6, p. 212-222.

VAN OOSTROM, S.H., J.R. ANEMA, B. TERLUIN, A. VENEMA, H.C.W. DE VET et W. VAN MECHELEN (2007). « Development of a workplace intervention for sick-listed employees with stress-related mental disorders : Intervention mapping as a useful tool », *BMC Health Services Research*, vol. 7, p. 127-127.

WALKER, L.O. et K.C. AVANT (2005). « Concept analysis », dans M. Connor (dir.), *Strategies for Theory Construction in Nursing*, 4e éd., Upper Saddle River, Pearson Education, p. 63-84.

WILSON, J. (1963). *Thinking with Cconcepts*, Cambridge, Cambridge University Press.

CHAPITRE 7

LA MÉTHODE DE LA REVUE SYSTÉMATIQUE
Illustration provenant du domaine de la toxicomanie et des troubles mentaux concomitants chez les jeunes

Karine Bertrand
Nadia L'Espérance
Jorge Flores Aranda

FORCES

- Les méthodes explicites de la revue systématique augmentent la validité interne de la démarche (p. ex., exhaustivité des sources d'information).
- Elle facilite un accès rapide à un bilan critique des connaissances scientifiques permettant d'orienter les prises de décision sur les plans des politiques de santé et de l'organisation des soins.
- Elle peut accélérer la mise en œuvre de stratégies thérapeutiques reposant sur des données probantes.

LIMITES

- Comparativement à une revue narrative, par exemple, elle exige beaucoup en fait de temps investi et de ressources financières et humaines.
- Elle n'est pas toujours appropriée, notamment lorsque les études dans un domaine particulier sont rares et hétérogènes.
- Lorsqu'une revue de littérature est menée pour définir une question de recherche novatrice, une stratégie de recherche plus inductive que la revue systématique peut être plus utile.

La revue de littérature, pouvant prendre diverses formes, est un moyen largement utilisé par une variété d'acteurs, notamment dans le domaine de la santé mentale. Le citoyen qui cherche à prendre une décision éclairée quant à sa propre santé aura parfois le réflexe de rechercher une revue de littérature qui saura le guider dans ses décisions. D'ailleurs, partant du besoin des personnes de se renseigner pour prendre des décisions éclairées, plusieurs organismes mettent à la disposition du public un bilan critique des connaissances disponibles dans un domaine de santé particulier. Citons par exemple le document rédigé par Éduc'alcool (2009) qui vulgarise un état des connaissances sur les effets de l'alcool sur le cerveau des adolescents.

Dans le monde de la santé, le courant des « meilleures pratiques », traduction de *best practices*, est largement répandu en Amérique du Nord, notamment dans le domaine du traitement des toxicomanies et des troubles mentaux concomitants. À titre d'exemple, notons les guides publiés par le Centre canadien de lutte contre l'alcoolisme et les toxicomanies (CCLAT, 2005), Santé Canada (2002a, 2002b), et les guides produits par l'Association des centres de réadaptation en dépendance du Québec (ACRDQ) (Bertrand *et al.*, 2006 ; Desrosiers et Jacques, 2009 ; Tremblay *et al.*, 2010). Au Québec, l'Institut national d'excellence en santé et services sociaux (INESSS), de par sa mission de promotion de l'excellence clinique, élabore des guides de pratiques qui s'appuient sur des revues de littérature rigoureusement menées, plus précisément des revues systématiques (INESSS, 2013). D'ailleurs, l'INESSS (2013), dans son guide sur la production de revues systématiques, souligne l'importance de cette méthodologie pour apporter des preuves scientifiques en vue d'aider à la prise de décision. Dans le même sens, plusieurs chercheurs publient des revues systématiques dont l'objectif est de conclure sur les traitements et pratiques basés sur des données probantes disponibles pour un sous-groupe clinique donné (Gowing, Ali et White, 2000 ; Lee et Rawson, 2008 ; Montag *et al.*, 2012).

La revue systématique constitue une méthode de choix pour formuler des recommandations de meilleures pratiques dans un domaine de santé particulier (Beauchamp et Duplantie, 2008 ; Harbour et Miller, 2001). Effectivement, les pratiques basées sur des résultats probants sont celles qui ont démontré leur efficacité à travers des études rigoureusement menées et répétées (Lundgren, Schilling et Peloquin, 2005,). En ce sens, une revue systématique permet d'établir dans quelle mesure plusieurs études contrôlées permettent de générer des résultats probants concernant l'efficacité d'une intervention ou d'un programme. Ainsi, lorsque plusieurs études rigoureusement menées en viennent aux mêmes conclusions, nous pouvons conclure à un résultat probant.

Avant d'aller plus loin, il importe de définir les concepts centraux ici en jeu, en précisant les différences entre deux grands types de revue de littérature, soit la revue narrative et la revue systématique. Par la suite, les principales étapes d'une revue systématique seront présentées. Enfin, le chapitre se termine par une illustration de ces différentes étapes, soit la description d'une démarche de revue systématique dans le domaine de la toxicomanie et des troubles mentaux concomitants chez les jeunes.

1. DÉFINITION DES CONCEPTS

1.1. Revue narrative

La revue narrative a pour but de regrouper de façon cohérente des études hétérogènes (Barnett-Page et Thomas, 2009). C'est une approche plutôt inductive qui met l'accent sur l'analyse de la méthodologie employée et sur la description des résultats des études recensées (Forbes et Griffiths, 2002). Elle permet de déceler des lacunes observées dans la littérature (Lucas *et al.*, 2007) et de constater les similarités et les divergences entre les études et leurs résultats (Boaz, Ashby et Young, 2002). La revue narrative permet donc de dresser un bilan des connaissances disponibles sur un sujet précis à partir de la littérature pertinente, sans reposer sur un processus méthodologique systématique, explicite, de recherche et d'analyse des articles inclus dans la revue (Horvath et Pewsner, 2004)

1.2. Revue systématique

Les revues systématiques constituent une activité de recherche en soi. Elles peuvent être considérées comme des études observationnelles rétrospectives concernant un niveau de preuve scientifique disponible pour un sujet donné (Egger et Smith, 2001; Horvath et Pewsner, 2004). La revue systématique a pour but de repérer, évaluer et synthétiser des données (Petticrew, 2001; Bambra, 2011; Egger et Smith, 2001).

Une bonne revue systématique doit présenter les caractéristiques suivantes: 1) une question de recherche bien définie ou une hypothèse à vérifier; 2) une démarche de recherche documentaire explicite permettant de repérer tous les documents pertinents; 3) l'évaluation de manière systématique de la qualité des méthodes employées dans les études recensées; 4) des conclusions fondées sur les résultats des études les plus rigoureuses

en termes méthodologiques (Petticrew, 2001). En outre, une revue systématique s'appuie sur une méthodologie clairement exposée et qui peut être reproduite par d'autres chercheurs (Greenhalgh, 1997; Bambra, 2011).

Des revues systématiques peuvent être menées sur des essais cliniques, des études observationnelles, des études cliniques, des études qualitatives, etc. En matière de méthodes d'analyses, les revues systématiques comportent parfois des analyses statistiques des données recueillies, et il s'agit alors d'une métaanalyse (voir le chapitre 8 de cet ouvrage). Les revues systématiques peuvent aussi reposer sur des analyses dites narratives ou thématiques (Petticrew, 2001). Ainsi, il peut y avoir des revues systématiques avec ou sans métaanalyse (Horvath et Pewsner, 2004; Leucht, Kisslinget et Davis, 2009; Egger, Smith et O'Rourke, 2001).

La principale différence entre une revue narrative et une revue systématique est l'application explicite et systématique d'une méthodologie rigoureuse. En outre, de façon générale, les revues narratives couvrent des sujets plus larges, tandis que les revues systématiques se concentrent sur des sujets plus ciblés, bien définis (Horvath et Pewsner, 2004) en plus d'évaluer systématiquement la qualité des études selon des critères précis et préétablis (Petticrew, 2001).

1.3. Étapes de la réalisation d'une revue systématique

La planification de toutes les étapes d'une revue systématique exige la rédaction d'un protocole. Dans celui-ci, il est nécessaire de formuler une question de recherche, d'établir les critères d'inclusion et d'exclusion des études, de définir les stratégies de recherche à employer (banques de données à consulter, choix de mots clés, instrument de mesure pour évaluer la qualité des publications recensées) ainsi que les méthodes d'analyses visées (Leucht *et al.*, 2009).

La composition de l'équipe de recherche qui mènera la revue systématique devrait tenir compte des expertises qui seront nécessaires par rapport tant au thème à l'étude qu'aux méthodes utilisées. Il est souvent conseillé d'inclure une personne ayant une expertise dans la conduite de revues systématiques ainsi qu'un spécialiste de la recherche documentaire. Enfin, lorsque le thème de la revue cible des pratiques cliniques, la formation d'un comité consultatif, formé d'informateurs clés du «terrain», est conseillée (Center for Reviews and Dissemination, 2009). Chacune des étapes décrites précédemment, résumées dans le tableau 7.1, sont expliquées en détail dans le texte qui suit.

Tableau 7.1.
Étapes d'une revue systématique

Étapes	Activités
1. Question de recherche et critères d'inclusion et d'exclusion	▪ Formuler la question de recherche : ▪ Définir le type de **p**articipants, l'**i**ntervention, l'intervention de **c**omparaison et les résultats (***o**utcomes*) : PICO ▪ Tenir compte du contexte et du devis méthodologique (***s**tudy design*) : PICO**S** ▪ Établir les critères d'inclusion et d'exclusion ▪ Définir de façon opérationnelle les principaux concepts ▪ Atteindre un équilibre entre critères trop ou pas assez précis
2. Recherche et sélection des études pertinentes	▪ Cibler la stratégie de recherche documentaire : ▪ Banques de données et mots clés ▪ Références secondaires ▪ Littérature grise ▪ Faire vérifier par au moins deux réviseurs le respect des critères d'inclusion et d'exclusion ▪ Recourir à un professionnel en documentation ▪ Établir une liste des documents non retenus et des raisons inhérentes
3. Évaluation de la qualité des études retenues	▪ Évaluer la méthodologie des études ▪ Choisir entre plusieurs grilles disponibles selon les devis, par exemple l'outil Cochrane (Higgins et Green, 2011) ▪ Faire effectuer l'analyse par deux chercheurs
4. Analyse et interprétation des résultats	▪ Répondre aux objectifs de départ ▪ Mettre en commun les évaluations des chercheurs et leurs conclusions : recherche de consensus ▪ Rédiger la synthèse narrative ▪ Trois processus : 1) Organiser la description des études en catégories logiques 2) Analyser les résultats pour chacune des catégories 3) Synthétiser les résultats de façon transversale, pour l'ensemble des études

1.3.1. Question de recherche et critères d'inclusion et d'exclusion

Afin de formuler la question de recherche, lorsque celle-ci est relative à l'efficacité d'une intervention, on peut répondre aux questions suivantes : « Qui sont les **p**articipants ? », « Quelle est l'**i**ntervention ? », « Quelle est l'intervention de **c**omparaison ? » et « Quels sont les résultats (***o**utcomes*) ? » (acronyme PICO) (Bambra, 2011 ; Grimshaw, 2010 ; INESSS, 2013). Pour

certains auteurs, il est nécessaire d'ajouter le contexte dans lequel les études ont été menées (Bambra, 2011) ainsi que la méthodologie (devis), ce qui ajoute un «S» à l'acronyme PICO (***s**tudy design*) (Center for Reviews and Dissemination, 2009). Notons toutefois que les revues systématiques dans le domaine de la santé portent parfois sur plusieurs interventions et que cette approche pour formuler la question de recherche n'est pas toujours complètement applicable. Par ailleurs, dans certains cas, les revues systématiques en santé ciblent une question de recherche moins spécifique que celles formulées à partir des composantes du PICO (Bambra, 2005). Il demeure que ces éléments peuvent guider le processus d'élaboration d'une question de recherche en santé.

La formulation claire d'une question de recherche permettra d'établir les critères d'inclusion et d'exclusion des études recensées. À cette étape, une définition opérationnelle des principaux concepts à l'étude est nécessaire. Ceci permettra aux chercheurs d'inclure les documents pertinents à analyser. Cependant, si les critères d'inclusion sont trop précis, cela risque d'exclure des études potentiellement intéressantes. D'un autre côté, si les critères sont trop larges, il sera difficile de comparer et de synthétiser les documents retenus (Center for Reviews and Dissemination, 2009). Certains auteurs recommandent de diviser une question large en plusieurs questions concises (Bambra, 2011).

1.3.2. Recherche et sélection des études pertinentes

Une fois la question de recherche définie, il est important de cibler la stratégie de recherche documentaire. Les sources de données doivent être décrites explicitement, par exemple : banques de données électroniques, listes des références citées dans les études retenues, revues clés du domaine étudié, sites Web d'organismes gouvernementaux, experts consultés, etc. Le choix des mots clés pour guider la recherche est important. Ceux-ci peuvent provenir de l'expertise des chercheurs concernant les termes courants utilisés dans leur discipline ainsi que des thesaurus des banques de données consultées (Hammerstrøm, Wade et Jorgensen, 2010). Les documents sont par la suite sélectionnés en fonction des critères d'inclusion et d'exclusion préalablement établis (Bambra, 2011). Il est conseillé qu'au moins deux réviseurs vérifient le respect de critères d'inclusion et d'exclusion afin d'éviter des biais de lecture (Leucht *et al.*, 2009 ; Egger et Smith, 2001). En outre, la réalisation d'une revue systématique par au moins deux personnes permet de partager une charge importante de travail. Aussi, le recours à un professionnel en documentation est fortement recommandé, notamment pour assurer l'exhaustivité de la recherche de sources documentaires

(Bambra, 2011; Center for Reviews and Dissemination, 2009). Enfin, une liste des études non retenues doit être établie et les raisons de l'exclusion doivent être documentées (Egger et Smith, 2001).

1.3.3. Évaluation de la qualité des études retenues

Les résultats des revues systématiques se fondent sur les résultats les plus rigoureux disponibles. Ainsi, l'évaluation de la qualité des études recensées doit tenir compte de la méthodologie employée dans celles-ci. Par exemple, plusieurs grilles d'analyses sont proposées pour l'évaluation systématique de la qualité et de la pertinence des études ciblées dans le cadre de revues systématiques, les grilles variant le plus souvent selon le type d'études recensées. Notons toutefois que celles portant sur les essais contrôlés randomisés sont plus nombreuses (Thomas et Harden, 2008. À titre d'exemple, l'échelle de Jadad (Jadad *et al.*, 1996) et l'outil Cochrane (Higgins et Green, 2011) sont largement utilisés dans le cadre de revues systématiques pour évaluer rigoureusement la qualité des études retenues. Il est aussi possible de mener une revue systématique d'études qualitatives, notamment en les évaluant avec une grille comme celle proposée par Mays et Popes (1995), ou encore une revue systématique de revues systématiques, celles-ci comparées et analysées avec la grille AMSTAR (Shea *et al.*, 2009). Il est recommandé que cette analyse se fasse parallèlement par deux chercheurs.

1.3.4. Analyse et interprétation des résultats

Les résultats d'une revue systématique doivent être interprétés et rapportés de façon à répondre aux objectifs de départ, que ce soit en matière de recherche, de développement de politiques ou d'interventions (Bambra, 2011). Par ailleurs, lors de la publication d'une revue systématique, il est important de décrire clairement les méthodes employées, les forces et limites des études recensées tout comme les forces et limites de la revue elle-même (Bambra, 2011). Lors de cette étape d'analyse et d'interprétation des résultats, les chercheurs qui ont mené parallèlement l'évaluation de la qualité des études croisent les résultats de leur analyse critique concernant la qualité de celles-ci, mais aussi leurs conclusions. Cet exercice de concordance permet d'en arriver à une meilleure validité des conclusions. Ainsi, plus concrètement, au moins deux chercheurs analysent de façon parallèle les études retenues puis discutent de leurs conclusions pour en arriver à un consensus, pour chacun des critères pris en compte par l'outil d'évaluation choisi. Advenant un désaccord, le recours à un troisième chercheur est recommandé (Center for Reviews and Dissemination, 2009). Ensuite, la rédaction des résultats de l'analyse critique, soit la synthèse narrative, comporte trois processus, décrits par Petticrew et Roberts (2006, p. 170):

1) organiser la description des études en catégories logiques ; 2) analyser les résultats pour chacune des catégories ; 3) synthétiser les résultats de façon transversale, pour l'ensemble des études.

Pour mieux comprendre ces trois processus nécessaires à la rédaction de la synthèse narrative, voyons quelques éléments concrets pour illustrer cette étape de la revue systématique. Ainsi, pour le premier processus, les catégorisations peuvent être multiples et dépendent de la question de recherche qui a guidé la revue systématique. Par exemple, les types d'intervention, les sous-groupes populationnels ou les différents indicateurs d'efficacité d'intérêt pourraient être des catégories. Les critères d'évaluation de qualité retenus sont d'autres exemples de sous-thèmes pertinents. Ce premier processus est particulièrement crucial lorsqu'une revue couvre plusieurs études. En effet, une description détaillée de chacune des études l'une à la suite de l'autre serait très difficile à suivre pour un lecteur et peu utile pour dégager les constats centraux à retenir. En ce qui concerne le deuxième processus, celui-ci consiste bien souvent, dans le cas des évaluations d'interventions, à rédiger une synthèse des résultats de chacune des études concernant des indicateurs de succès clés en plus d'une description de la qualité des études. Le recours à un tableau pour synthétiser les informations est recommandé. Pour ce qui est du troisième processus, soit la synthèse de l'analyse transversale de l'ensemble des études, celle-ci permet de situer le lecteur concernant les connaissances disponibles globalement et les principaux résultats. Bien que ce processus d'analyse se réalise plus tard dans la démarche, il permet de rédiger le début de la section des résultats de la revue systématique. Par exemple, une revue commence typiquement par une description de l'ampleur des connaissances sur le sujet ciblé, comme le nombre d'études correspondant aux critères d'inclusion, les types de devis et de contextes, et les caractéristiques des populations étudiées. Aussi, les principaux résultats liés à la question de recherche, typiquement l'efficacité d'interventions dans un domaine de santé particulier, sont résumés d'entrée de jeu. Par la suite, la description de chacune des études peut être réalisée, selon les catégorisations choisies, en s'appuyant sur un tableau-synthèse.

2. ILLUSTRATION D'UNE REVUE SYSTÉMATIQUE DANS LE DOMAINE DE LA TOXICOMANIE ET DES TROUBLES MENTAUX CONCOMITANTS CHEZ LES JEUNES

Cette deuxième section du chapitre vise à illustrer de façon concrète les quatre grandes étapes pour mener une revue systématique ayant été précédemment décrites. Cette illustration provient d'une revue systématique ayant été menée sur le traitement de la toxicomanie et des troubles mentaux concomitants chez les jeunes.

2.1. Question de recherche et critères d'inclusion et d'exclusion

La question de recherche de cette revue systématique s'inscrit dans la suite d'une première démarche réalisée en 2006, soit la rédaction d'un guide de meilleures pratiques sur le traitement de la toxicomanie chez les adolescents (Bertrand *et al.*, 2006). Au moment de la rédaction du guide en 2006, très peu de connaissances étaient disponibles concernant l'efficacité du traitement des troubles concomitants chez les jeunes. Pourtant, tant dans les populations cliniques (Chan, Dennis et Funk, 2008) que dans la population générale (Merikangas *et al.*, 2010), les taux élevés de comorbidité entre les troubles liés aux substances psychoactives et les troubles mentaux chez les jeunes sont bien documentés. Ainsi, les cliniciens œuvrant dans les milieux jeunesse, que ce soit dans le domaine de la santé mentale ou celui de la toxicomanie, sont nombreux à nommer le besoin d'outils thérapeutiques et de formation sur le plan des troubles concomitants. D'un point de vue tant clinique que scientifique, le repérage de données probantes en ce domaine répond clairement à des préoccupations actuelles pour arriver à optimiser l'efficacité des interventions auprès d'un sous-groupe d'adolescents particulièrement vulnérable. Ainsi, la question de recherche à l'origine de cette revue systématique est la suivante : « Quels sont les modèles intégrés d'intervention psychosociale de la toxicomanie et des troubles mentaux concomitants chez les adolescents dont l'efficacité repose sur des données probantes ? »

Soulignons que la formulation de la question a été guidée par l'acronyme PICOS discuté précédemment, en vue d'en assurer un niveau adéquat de spécificité : définir le type de **p**articipants, l'**i**ntervention, l'intervention de **c**omparaison, les résultats (***o****utcomes*) et le contexte ainsi que le devis (***s****tudy design*). Les critères d'inclusion discutés ultérieurement ont permis de préciser davantage ou de compléter la prise en compte des dimensions du PICOS. Soulignons toutefois que compte tenu des connaissances limitées dans le domaine ciblé, il a été convenu de ne pas limiter la question de recherche selon le type d'intervention de comparaison considéré.

Le choix de la revue systématique de données probantes s'appuie sur le fait que cette stratégie de recherche est recommandée dans le processus d'élaboration de guides de meilleures pratiques (INESSS, 2013 ; Harbour et Miller, 2001). Cette démarche constitue une première étape dans la mise à jour de nos connaissances pour en arriver à dégager des constats utiles pour la pratique. En effet, sachant que le domaine de recherche d'intérêt demeure assez récent et encore trop peu exploré, il est attendu que les résultats de cette revue systématique appuieront la nécessité de poursuivre d'autres démarches pour répondre aux questionnements des cliniciens. Ainsi, à l'instar de ce qui s'était produit lors de la démarche pour développer

les guides de meilleures pratiques en toxicomanie (Bertrand *et al.*, 2006 ; Tremblay *et al.*, 2010), nous devrons ensuite élargir notre recherche documentaire au-delà des données probantes. Il faudra donc recenser des travaux reposant sur divers devis de recherche. En complément, il sera également important de mettre sur pied des groupes d'experts en vue de dégager des consensus permettant d'éclairer des enjeux cliniques auxquels les connaissances actuelles ne permettent pas de répondre. Tout de même, dans un premier temps, il nous faut d'abord connaître avec précision les données probantes permettant d'appuyer l'efficacité de modèles d'intervention ciblant les troubles concomitants chez les jeunes.

Pour répondre spécifiquement à la question de recherche, nous avons donc établi les critères d'inclusion suivants, qui permettront de sélectionner les articles scientifiques pertinents : 1) le devis de recherche doit reposer sur un essai randomisé ; 2) le traitement expérimental dont on évalue l'efficacité doit comprendre une dimension psychosociale et cibler spécifiquement un trouble lié aux substances psychoactives et un autre trouble mental concomitant ; 3) les participants doivent être aux prises avec un trouble lié aux substances et un trouble mental spécifique concomitant, tel qu'évalué au début du traitement ; 4) les participants doivent être âgés de 13 à 21 ans ; 5) l'étude doit avoir été publiée au cours des vingt dernières années, donc après 1992 ; 6) les publications doivent être en français ou en anglais.

Nous avons retenu les essais randomisés, puisqu'il s'agit du type de devis dont la rigueur de la méthodologie permet de conclure à des liens de cause à effet entre un traitement et ses effets thérapeutiques (voir le chapitre 10 de cet ouvrage). Le niveau de preuve le plus élevé concernant l'efficacité d'un traitement, alors défini comme étant basé sur des données probantes, provient d'un cumul d'essais randomisés qui en viennent aux mêmes conclusions. Le modèle de traitement évalué par les études retenues doit cibler à la fois un trouble lié aux substances et un trouble mental concomitant, pour être retenu dans la revue systématique. Les études documentant les effets d'un traitement de la toxicomanie en général sur les troubles concomitants ou encore l'inverse sont exclus, car ceux-ci ne correspondent pas aux caractéristiques d'un traitement adapté aux troubles concomitants. De même, les essais randomisés s'attardant à l'efficacité d'un médicament, sans inclure une composante psychosociale dans le traitement, n'ont pas été retenus. La période de l'adolescence a été ciblée, car elle comporte des enjeux cliniques spécifiques, différents de ceux des jeunes adultes. Comme l'âge limite de la majorité est de 21 ans dans plusieurs endroits en Amérique du Nord et que ce groupe d'âge est souvent inclus dans les services jeunesse en toxicomanie au Québec, c'est la limite d'âge qui a été déterminée pour la revue systématique. Enfin, les essais randomisés portant sur des populations

cliniques hétérogènes dont la majorité sont aux prises avec divers troubles concomitants, mais pas nécessairement tous, ont été exclus. En effet, ce type d'études ne permet pas de répondre à la question de recherche, car les conclusions ne peuvent permettent d'établir l'efficacité d'un modèle d'intervention intégré pour un sous-groupe spécifique.

2.2. Recherche et sélection des études pertinentes

La recherche documentaire a été effectuée en février 2013 à partir des principales banques de données pertinentes pour le domaine de recherche : Medline, Psychinfo, SocIndex, Social work abstract, ERIC. La stratégie de recherche documentaire a été élaborée par les chercheurs avec l'appui d'une bibliothécaire de l'Université de Sherbrooke (Québec). Le recours aux thesaurus des banques de données ainsi que les connaissances des chercheurs concernant les termes propres à la discipline ont permis de raffiner la stratégie de recherche documentaire. Les mots clés suivants ont été utilisés pour chacun des concepts : 1) le trouble lié aux substances : « *addiction OR drug abuse OR alcohol abuse OR substance abuse OR substance related disorders* » ; 2) la comorbidité : « *comorbid* OR dual diagnosis OR dual disorders OR co-occuring* » 3) jeune : « *youth OR adolescen* OR teen* OR juvenile* » ; 4) le traitement : « *treatment OR psychotherapy OR intervention OR rehabilitation* » ; 5) le devis de l'étude : « *Randomi?ed*[2] *controlled trials OR RCT OR experimental methods OR experimental designs OR clinical trials OR random allocation* ». Tous les résultats de recherche pour chacun de ces concepts ont été croisés au moyen de l'opérateur booléen « ET ». Un premier groupe d'articles pertinents à la question de recherche a été repéré à partir de la lecture de la liste des titres et résumés d'études. À la suite de cette première étape de recherche, les listes de références des documents pertinents repérés ont été consultées, notamment celles des revues de littérature sur le sujet ayant été dégagées par la recherche documentaire. Quelques experts ont été consultés pour compléter la recherche documentaire, soit un gestionnaire responsable de la mise à jour des programmes de traitement en dépendance pour les jeunes dans son établissement, une chercheuse menant des travaux de recherche auprès d'adolescents en traitement de la toxicomanie, et deux cliniciennes impliquées dans un de nos projets de recherche auprès de ce sous-groupe clinique ainsi que dans des comités de travail sur le sujet. De plus, la consultation de la littérature grise a été effectuée, soit plus précisément les documents d'orientation et les publications sur des sites Web d'organismes reconnus, comme ceux

2. Le symbole « ? » permet de remplacer la lettre « s » ou « z » lors de la recherche dans les banques de données, puisque les deux types d'orthographe sont acceptées (*randomised*/*randomized*), selon la langue choisie, soit l'anglais américain ou britannique.

du SAMHSA (Substance Abuse and Mental Health Services Administration), du CAMH (Centre for Addiction and Mental Health) et du CCLAT (Centre canadien de lutte contre l'alcoolisme et les toxicomanies). La sélection finale des études retenues s'est faite à partir de la lecture des documents pertinents et à la suite d'un consensus par les chercheurs.

De façon plus précise, le tableau 7.2 résume les résultats de la recherche documentaire réalisée, en respectant l'ordre de consultation de chacune des sources de données. Au total, 141 sources documentaires ont été repérées, dont quatre ont été sélectionnées pour la revue systématique. Les principales raisons d'exclusion concernent le devis d'étude autre qu'un essai randomisé (*n* = 88/137 références exclues) ainsi que la nature du modèle de traitement évalué, soit non intégré (*n* = 24/137 références exclues). En effet, plusieurs études portaient sur un modèle de traitement ciblant uniquement l'une des problématiques, soit un trouble lié aux substances, soit un trouble mental.

Tableau 7.2.
Synthèse des résultats de la recherche documentaire dans les banques de données

Sources consultées (dans l'ordre)	**Références obtenues (*n*)**	**Nouvelles références (*n*)**	**Références retenues (*n*)**	**Raisons des exclusions (*n*)**
1. PsycINFO	57	57	1	N'est pas un essai randomisé (35). N'évalue pas un traitement intégré (14). Ne vise pas la population cible (5). Autres raisons (2).
2. Medline	81	66	1	N'est pas un essai randomisé (44). N'évalue pas un traitement intégré (5). Ne vise pas la population cible (10). Autres raisons (6).
3. SocIndex	22	14	0	N'est pas un essai randomisé (8). N'évalue pas un traitement intégré (4). Ne vise pas la population cible (2).
4. Social work abstracts	3	1	0	N'est pas un essai randomisé (1).
5. ERIC	3	0	S. O.	S. O.
6. Liste de références des sources pertinentes	S. O.[a]	3	2	N'évalue pas un traitement intégré (1).
TOTAL	S. O.	141	4	137 références exclues au total.

a : Pour les sources documentaires repérées autrement que par l'intermédiaire des banques de données, seules les nouvelles références ont été recensées systématiquement

2.3. Évaluation de la qualité des études retenues

L'analyse de la qualité des études a été menée de façon systématique pour les quatre études sélectionnées (Azrin *et al.*, 2001 ; Najavits, Gallop et Weiss, 2006 ; Riggs *et al.*, 2007 ; Thurstone *et al.*, 2010). L'outil d'évaluation de la qualité retenu est celui recommandé dans le *Cochrane Handbook for Systematic Reviews of Intervention* (<http://www.handbook.cochrane.org> ; Higgins et Green, 2011). Cet outil évalue la qualité des études selon six dimensions : 1) la méthode pour générer la séquence de randomisation (biais de sélection) ; 2) la dissimulation de la séquence de randomisation (biais de sélection) ; 3) la méthode utilisée pour que les participants et le personnel de recherche soient à l'insu (biais de performance) ; 4) la méthode pour que l'évaluateur aux suivis soit à l'insu (biais de détection) ; 5) l'ampleur des données manquantes (biais d'attrition) ; 6) les résultats rapportés de façon complète et non sélective (*reporting bias*). L'outil Cochrane décrit également des critères précis pour appuyer l'évaluation de chacune de ces dimensions. Toute autre limite méthodologique est aussi notée de façon systématique, à la suite de l'évaluation des six dimensions. Le tableau 7.3 illustre la synthèse de l'analyse de la qualité selon ces six critères pour l'ensemble des quatre études sélectionnées. Le tableau 7.4 permet d'illustrer le résumé de l'évaluation de la qualité d'une seule de ces quatre études, à titre d'exemple d'application de l'outil Cochrane.

Tableau 7.3.
Sommaire de l'évaluation de la qualité des études sélectionnées selon les risques de biais, évalués au moyen de l'outil Cochrane

	Azrin *et al.*, 2001	**Najavits *et al.*, 2006**	**Riggs *et al.*, 2007**	**Thurstone *et al.*, 2010**
Séquence de randomisation générée au hasard (biais de sélection)	?*	?	?	+
Dissimulation de la séquence de randomisation (biais de sélection)	–	+	+	+
L'intervention : participants et personnel aveugles (biais de performance)	+	?	+	+
Évaluateur aux suivis : aveugle (biais de détection)	+	+	+	+
Ampleur des données manquantes (biais d'attrition)	–	?	+	+
Les résultats rapportés de façon complète et non sélective (*reporting bias*)	+	+	+	+

* (+) = faible risque ; (?) = risque incertain ; (–) = risque élevé.

Tableau 7.4.
Exemple d'une grille d'évaluation basée sur l'outil Cochrane de la qualité de l'étude de Riggs *et al*. (2007)

Dimension évaluée (biais)	Jugement (niveau de risque)	Justification du jugement
La méthode pour générer la séquence de randomisation (biais de sélection)	Risque incertain	▪ La méthode de randomisation par bloc utilisée est adéquate, mais il n'y a aucun détail sur le processus ayant permis de générer ces blocs, ce qui ne permet pas de juger un « faible risque ».
La dissimulation de la séquence de randomisation (biais de sélection)	Faible risque	▪ Procédure pertinente décrite en détail pour assurer le double insu chercheurs/participants lors de la répartition des participants entre les deux groupes.
La méthode pour que les participants et le personnel de recherche soient à l'insu (biais de performance)	Faible risque	▪ Les participants et les cliniciens de recherche n'ont pas été informés de la nature de la médication prescrite (fluoxetine ou placebo) durant toute la durée de l'étude.
La méthode pour que l'évaluateur soit à l'insu lors des suivis (biais de détection)	Faible risque	▪ Le personnel de recherche qui mesure les résultats de l'intervention et le degré d'adhésion au traitement n'a pas été informé de la nature de la médication utilisée (fluoxetine ou placebo) durant toute la durée de l'étude.
L'ampleur des données manquantes (biais d'attrition)	Faible risque	▪ Faible abandon dans les deux groupes comparés : 52/63 et 54/63 participants ont terminé l'étude. ▪ Tous les patients randomisés ont été inclus dans les analyses. ▪ On connaît le cheminement conduisant à l'échantillon final, la proportion des participants exclus ou n'ayant pas terminé l'étude, ainsi que les caractéristiques sociodémographiques de l'échantillon.
Les résultats rapportés de façon complète et non sélective (*reporting bias*)	Faible risque	▪ Les résultats pour chaque mesure de succès thérapeutique sont présentés en détail, pour chaque temps de mesure (semaines 0, 4, 8, 12 et 16). ▪ L'absence de résultats statistiquement significatifs est soulignée systématiquement. ▪ Les auteurs présentent aussi leur analyse post hoc, en les différenciant des analyses prévues initialement dans le protocole.
Autre biais ou limite	S. O.	▪ L'applicabilité des résultats dans un contexte habituel de soins est questionnable : la visite hebdomadaire de l'infirmière est une procédure de recherche visant à améliorer l'adhésion au traitement et les taux de suivi de l'étude ; cette mesure peut cependant contribuer à l'explication des résultats thérapeutiques sans qu'on puisse estimer l'ampleur de cet effet.

2.4. Analyse et interprétation des résultats

La revue systématique de la littérature a permis de retenir seulement quatre études randomisées (Azrin *et al.*, 2001; Najavits *et al.*, 2006; Riggs *et al.*, 2007; Thurstone *et al.*, 2010) qui correspondent aux critères d'inclusion et qui permettent de répondre à la question de recherche initiale. Ainsi, force est de constater les lacunes des connaissances concernant l'efficacité de traitements psychosociaux intégrés qui ciblent les troubles concomitants chez les adolescents. Comme illustré dans le tableau 7.5, les modèles d'intervention psychosociaux intégrés ayant fait l'objet d'une évaluation empirique rigoureuse sont rares et concernent une variété de troubles mentaux concomitants à la toxicomanie: les troubles de la conduite, le syndrome de stress post-traumatique (SSPT), la dépression majeure et le trouble de déficit de l'attention avec hyperactivité (TDAH). Tout de même, il faut souligner que toutes les études sélectionnées documentent des progrès significatifs des adolescents concernant les indicateurs de résultat du trouble lié aux substances (TS) et du trouble concomitant ciblé.

Toutefois, seul l'essai randomisé portant sur le modèle Seeking Safety (SS) permet de conclure à une supériorité d'un modèle sur un autre (Najavits *et al.*, 2006). Par ailleurs, comme cette conclusion ne s'appuie que sur une seule étude, il n'est pas possible d'en tirer des conclusions claires, le modèle SS ne pouvant être considéré comme basé sur des données probantes pour le moment. Dans l'étude de Riggs *et al.* (2007), l'ajout d'un médicament (fluoxetine) à la thérapie cognitive comportementale (TCC), comparativement à l'ajout d'un placebo, est plus efficace pour diminuer les symptômes dépressifs, selon le Childhood Depression Rating Scale–Revised. Toutefois, au sujet de l'abus de substances et des problèmes de comportement, aucune différence entre les groupes n'est constatée. Ainsi, dans l'ensemble, et ce, pour la majorité des études retenues, les jeunes ayant reçu un traitement vont mieux aux temps de suivi. Toutefois, le plus fréquemment, les résultats thérapeutiques favorables observés sont similaires, peu importe le type d'intervention reçue.

Ces résultats similaires entre les études pourraient s'expliquer par le fait que, malgré la grande diversité des types de troubles concomitants ciblés, les types d'interventions évalués sont similaires. Effectivement, toutes les études évaluent au moins un traitement manualisé, avec une composante cognitive comportementale, des cibles claires qui prennent en compte la comorbidité, tout en assurant une durée de traitement moyenne (12 à 25 sessions). Également, dans les quatre études retenues, la formation et la supervision des cliniciens offertes de façon soutenue et adéquate ainsi que des mesures pour assurer la qualité et l'intégrité du traitement offert sont détaillées. Ces modèles d'intervention correspondent donc aux

critères d'une intervention « *bona fide* » ; ces types d'intervention se sont largement montrés efficaces et équivalents dans divers domaines psychosociaux (Tremblay *et al.*, 2010). D'ailleurs, il n'est pas étonnant que le seul essai randomisé inclus dans cette revue systématique concluant à la supériorité d'un modèle compare un modèle « *bona fide* » (le SS) à un traitement habituel non standardisé. Ainsi, le type d'approche et les modalités du traitement habituel ne sont pas précisés, et les chercheurs ne décrivent pas les formations et supervisions qui auraient pu être offertes au groupe de cliniciens de ce groupe de comparaison (Najavits *et al.*, 2006).

Par ailleurs, malgré le peu de connaissances disponibles, il faut souligner la rigueur méthodologique, dans l'ensemble, des essais randomisés retenus par cette revue systématique. En effet, l'analyse de la qualité des études à partir de l'outil Cochrane permet de constater, pour la plupart des critères d'évaluation, des risques de biais qui sont faibles (voir le tableau 7.3). Le manque d'information concernant les procédures de randomisation constitue une lacune fréquente dans les études, ce qui peut soulever un certain doute quant au risque lié au biais de sélection. Dans cette veine, la procédure de randomisation de l'étude d'Azrin *et al.* (2001), laquelle s'appuie notamment sur un jugement clinique en vue de former des paires d'adolescents possédant des caractéristiques cliniques similaires, peut nous faire douter de la rigueur du processus. D'ailleurs, cette même étude comporte aussi un biais d'abandon important, puisque les résultats de l'essai randomisé portent uniquement sur les 56 adolescents ayant suivi huit sessions sur 15, parmi les 88 jeunes ayant entamé la thérapie. Les caractéristiques des jeunes qui persistent dans un traitement peuvent être fort différentes de celles liées aux adolescents qui abandonnent prématurément leur démarche.

D'autres limites des quatre études retenues concernent la « transférabilité » des résultats. D'abord, toutes ces études se sont déroulées dans des contextes de recherche difficilement reproductibles dans des contextes cliniques habituels. Aussi, les échantillons sont majoritairement, ou exclusivement dans un cas, constitués soit de filles, soit de garçons. Les enjeux cliniques concernant la différenciation du traitement selon le genre ne sont pas abordés. Enfin, ces essais randomisés ont été menés sur de petits échantillons, et les temps de suivis considérés sont relativement courts (quatre semaines à six mois), ce qui ne permet pas de comprendre la complexité du processus de rétablissement des adolescents qui y participent.

Tableau 7.5.
Quatre essais randomisés sélectionnés : description des études et résultats

Auteurs	**Population**	**Intervention et comparateur**	**Contexte (*setting*)**	**Résultats (*outcomes*)**
Azrin *et al.*, 2001	▪ Double diagnostic : TS (1) et trouble de la conduite ▪ *n* = 56 au total ▪ 46 gars/10 filles ▪ 12-17 ans ▪ Parent accepte de s'impliquer. ▪ Ont suivi au moins 8 sur 15 sessions	▪ Thérapie comportementale familiale (*n* = 29) ▪ Thérapie individuelle axée sur la résolution de conflits (*n* = 27) ▪ 2 groupes : 15 sessions/6 mois ▪ Pas d'hypothèse de supériorité d'un modèle ▪ Formation, supervision et contrôle de la qualité adéquats	▪ Références à la recherche : Centre de détention, juge, agent de probation, école ▪ Thérapeutes : étudiants en psychologie supervisés ▪ Suivis : pré-post-6 mois	▪ Principaux indicateurs : consommation de substances et problèmes de conduite ▪ Plusieurs autres indicateurs mesurés ▪ Les deux groupes sont équivalents : s'améliorent sur tous les indicateurs
Najavits *et al.*, 2006	▪ Double diagnostic : syndrome de stress post-traumatique et TS ▪ *n* = 33 ▪ Toutes des adolescentes ▪ 16 ans en moyenne	▪ SS (2) + TH (3) (*n* = 18) ▪ TH seul (*n* = 15) ▪ SS : 25 sessions/3mois ▪ TH : durée et type variables ▪ Formation, supervision et contrôle de la qualité adéquats, mais pour SS seulement	▪ SS offert dans un contexte de recherche standardisé ▪ Recrutement par l'équipe de recherche par l'intermédiaire d'affiches, de visites dans les cliniques, etc. ▪ Suivis pré- et post-3 mois	▪ Le groupe SS s'améliore davantage que le groupe de comparaison sur des indicateurs d'abus de substances et de trauma + sur d'autres dimensions
Riggs *et al.*, 2007	▪ Double diagnostic : dépression majeure et TS ▪ *n* = 126 ▪ 85 hommes/41 femmes ▪ Âge : 13-19 ans	▪ TCC (4) + fluoxetine (*n* = 63) ▪ TCC + placebo (*n* = 63) ▪ Durée de 16 semaines pour les 2 groupes ▪ Formation, supervision et contrôle de la qualité adéquats pour les deux groupes	▪ Recrutement dans la population générale et auprès de services sociaux et de justice juvénile ▪ TCC et fluoxetine offerts dans un contexte de recherche standardisé ▪ Suivis 0, 4, 8, 12 et 16 semaines	▪ TCC + fluoxetine : amélioration plus marquée que groupe de comparaison sur l'un des deux indicateurs de dépression ▪ 2 groupes s'améliorent sur divers indicateurs de TS et de troubles de conduite, de façon équivalente

(*suite*)

Tableau 7.5.
Quatre essais randomisés sélectionnés : description des études et résultats (*suite*)

Auteurs	**Population**	**Intervention et comparateur**	**Contexte (*setting*)**	**Résultats (*outcomes*)**
Thurstone *et al.*, 2010	▪ Double diagnostic : ▪ TS et TDAH (5) ▪ *n* = 70 ▪ 15 filles/55 gars ▪ 13 à 19 ans	▪ EM (6) + TCC + atomoxetine ▪ EM + TCC + placebo ▪ Durée : 12 semaines/sessions pour les 2 groupes ▪ Formation, supervision et contrôle de la qualité adéquats pour les deux groupes	▪ Traitements offerts dans un contexte de recherche standardisé ▪ Stratégies de recrutement non décrites ▪ Suivis 0, 4, 8, 12 semaines	▪ 2 groupes s'améliorent sur les indicateurs de TDAH et de TS

1 : TS = trouble lié aux substances.
2 : SS = Seeking Safety.
3 : TH = traitement habituel.
4 : TCC = thérapie cognitive-comportementale.
5 : TDAH = trouble de déficit de l'attention avec hyperactivité.
6 : EM = entretien motivationnel.

En somme, cet exemple de revue systématique permet partiellement de répondre à la question de recherche initiale, mais souligne surtout des lacunes importantes concernant les connaissances sur les traitements efficaces qui ciblent des troubles concomitants chez les adolescents. Cette revue systématique, décrite pour illustrer l'application de la méthode à un questionnement concret, est nécessaire, mais bien sûr insuffisante pour éclairer les cliniciens concernés par la problématique de l'étude.

Tout de même, les résultats de la revue systématique permettent de mettre en lumière certains ingrédients des modèles d'intervention psychosociaux qui semblent efficaces et viennent donc confirmer la pertinence des modèles d'approche cognitive-comportementale en ce domaine. Il est certainement nécessaire de mener d'autres essais randomisés pour confirmer l'efficacité de ces modèles d'intervention qui suivent une approche cognitive-comportementale. En effet, cette approche est très large et peut regrouper des modèles de traitement fort différents, selon par exemple les composantes de cette approche qui sont plus ciblées. Le développement et l'évaluation de modèles d'intervention intégrés ciblant précisément deux troubles concomitants constituent également une piste de recherche importante pour mieux répondre aux besoins des jeunes.

CONCLUSION

La revue systématique constitue une méthode de recherche largement utilisée dans le domaine de la santé mentale comme dans celui de la toxicomanie, particulièrement utile pour guider l'élaboration de guides de « meilleures pratiques ». L'utilisation de cette méthode constitue une étape essentielle permettant de révéler les pratiques cliniques basées sur des données probantes. La rigueur de la revue systématique est incontournable lorsqu'il est question de soutenir des décisions qui orientent une offre de soins auprès de clientèles vulnérables aux prises avec des besoins complexes. Par ailleurs, la revue systématique permet de cibler des domaines de recherche pour lesquels il faut mener de nouvelles études, comme illustré par l'exemple présenté dans ce chapitre. En ce sens, la revue systématique, bien qu'étant la méthode de revue de littérature la plus rigoureuse, n'est pas une panacée et doit être utilisée en complément à d'autres méthodes. Pour l'élaboration de guides de « meilleures pratiques », la revue systématique s'accompagne bien souvent de panels d'experts, par exemple. Les consensus qui se dégagent de l'analyse des opinions d'experts, combinée aux résultats d'une revue systématique, favorisent l'émergence de pratiques novatrices et adaptées aux défis actuels rencontrés par les décideurs politiques, les gestionnaires et les cliniciens.

RÉFÉRENCES

AZRIN, N.H., B. DONOHUE, G.A. TEICHNER, T. CRUM, J. HOWELLET et L.A. DECATO (2001). « A controlled evaluation and description of individual-cognitive problem solving and family-behavior therapies in dually-diagnosed conduct-disordered and substance-dependent youth », *Journal of Child & Adolescent Substance Abuse*, vol. 11, nº 1, p. 1-44.

BAMBRA, C. (2005). « Reviewing the evidence : Reflections from experience », *Evidence Policy : A Journal of Research, Debate and Practice*, vol. 1, nº 2, p. 243-256.

BAMBRA, C. (2011). « Real world reviews : A beginner's guide to undertaking systematic reviews of public health policy interventions », *Journal of Epidemiology and Community Health*, vol. 65, p. 14-19.

BARNETT-PAGE, E. et J. THOMAS (2009). « Methods for the synthesis of qualitative research : A critical review », *Medical Research Methodology*, vol. 9, nº 59, p. 1-11.

BEAUCHAMP, S. et J. DUPLANTIE (2008). *Vers une méthode d'élaboration des guides de pratique dans le secteur des services sociaux. Position du comité sur les guides de pratique en services sociaux de l'INESSS*, Montréal, Institut national d'excellence en santé et en services sociaux.

BERTRAND, K., C. BEAUMONT, L. MASSICOTTE et H. BLOUIN (2006). *Intervenir auprès des jeunes et de leur entourage dans les centres de réadaptation pour personnes alcooliques et toxicomanes. Pratiques gagnantes et offre de services de base*, Montréal, Fédération québécoise des centres de réadaptation pour personnes alcooliques et autres toxicomanes.

BOAZ, A., D. ASHBY et K. YOUNG (2002). *Systematic Reviews : What Have They Got to Offer Evidence Based Policy and Practice ?*, Londres, ESRC UK Centre for Evidence Based Policy and Practice Queen Mary, University of London.

CENTER FOR REVIEWS AND DISSEMINATION (2009). *Systematic Reviews. CRD's Guidance for Undertaking Reviews in Health Care*, Layerthorpe, York Publishing Services.

CENTRE CANADIEN DE LUTTE CONTRE L'ALCOOLISME ET LES TOXICOMANIES – CCLAT (2005). *Toxicomanie au Canada. Enjeux et options actuels*, Ottawa, Centre canadien de lutte contre l'alcoolisme et les toxicomanies.

CHAN, Y., M.L. DENNIS et R.R. FUNK (2008). « Prevalence and comorbidity of major internalizing and externalizing problems among adolescents and adults presenting to substance abuse treatment », *Journal of Substance Abuse Treatment*, vol. 34, nº 1, p. 14-24.

CLEARY, M., G.E. HUNT, S. MATHESON et G. WALTER (2008). « The association between substance use and the needs of patients with psychiatric disorder, levels of anxiety, and caregiving burden », *Archives of Psychiatric Nursing*, vol. 22, nº 6, p. 375-385.

DESROSIERS, P. et C. JACQUES (2009). *Les services en jeu pathologique dans les centres de réadaptation en dépendance. Guide de bonnes pratiques et offre de services de base*, Montréal, Association des centres de réadaptation en dépendance du Québec.

ÉDUC'ALCOOL (2009). *Les effets de la consommation précoce d'alcool. Causes et conséquences de la surconsommation à l'adolescence*, Montréal, Éduc'alcool.

EGGER, M. et D.G. SMITH (2001). « Principles of and procedures for systematic reviews », dans M. Egger, D.G. Smith et D.G. Altman (dir.), *Systematic Reviews in Health Care : Meta-Analysis in Context*, Londres, BMJ Books, p. 23-42.

EGGER, M., G.D. SMITH et K. O'ROURKE (2001). « Rationale, potentials, and promise of systematic reviews », dans M. Egger, G.D. Smith et D.G. Altman (dir.), *Systematic Reviews in Health Care : Meta-Analysis in Context*, Londres, BMJ Publishing, p. 3-19.

FORBES, A. et P. GRIFFITHS (2002). « Methodological strategies for the identification and synthesis of "evidence" to support decision-making in relation to complex healthcare systems and practices », *Nursing Inquiry*, vol. 9, nº 3, p. 141-155.

GOWING, L.R., R.L. ALI et J.M. WHITE (2000). « Systematic review processes and the management of opioid withdrawal », *Australian and New Zealand Journal of Public Health*, vol. 24, nº 4, p. 427-431.

GREENHALGH, T. (1997). « Papers that summarise other papers (systematic reviews and meta-analyses) », *British Medical Journal*, vol. 315, nº 7109, p. 672-675.

GRIMSHAW, J. (2010). *Guide sur la synthèse des connaissances*, <http://www.cihr-irsc.gc.ca/f/41382.html>, consulté le 12 janvier 2013.

HAMMERSTRØM, K., A. WADE et A. JORGENSEN (2010). *Searching for Studies : A Guide to Information Retrieval for Campbell Systematic Reviews*, Oslo, Campbell Systematic Reviews.

HARBOUR, R. et J. MILLER (2001). « A new system for grading recommendations in evidence based guidelines », *British Medical Journal*, vol. 323, nº 7308, p. 334-336.

HIGGINS, J. et S. GREEN (2011). *Cochrane Handbook for Systematic Reviews of Interventions*, version 5.1.0, Chichester, John Wiley & Sons.

HORVATH, A.R. et D. PEWSNER (2004). « Systematic reviews in laboratory medicine : Principles, processes and practical considerations », *Clinica Chimica Acta*, vol. 342, p. 23-39.

INSTITUT NATIONAL D'EXCELLENCE EN SANTÉ ET EN SERVICES SOCIAUX – INESSS (2013). *Les normes de production des revues systématiques. Guide méthodologique*, document rédigé par Valérie Martin et Jolianne Renaud sous la direction de Pierre Dagenais, Montréal, INESSS.

JADAD, A.R., A. MOORE, D. CARROLL, C. JENKINSON, J.M. REYNOLDS, D.J. GAVAGHAN *et al*. (1996). « Assessing the quality of reports of randomized clinical trials : Is blinding necessary ? », *Controlled Clinical Trials*, vol. 17, p. 1-12.

LEE, N.K. et R.A. RAWSON (2008). « A systematic of review of cognitive and behavioural therapies for methamphetamine dependence », *Drug and Alcohol Review*, vol. 27, p. 309-317.

LEUCHT, S., W. KISSLINGET et J.M. DAVIS (2009). « How to read and understand and use systematic reviews and meta-analyses », *Acta Psychiatrica Scandinavica*, vol. 119, p. 443-450.

LUCAS, P.J., J. BAIRD, L. ARAI, C. LAW et H.M. ROBERTS (2007). « Worked examples of alternative methods for the synthesis of qualitative and quantitative research in systematic reviews », *BMC Medical Research Methodology*, vol. 7, nº 4, p. 11.

LUNDGREN, L.M., R.F. SCHILLING et S.D. PELOQUIN (2005). « Evidence-based drug treatment practice and the child welfare system : The example of methadone », *Social Work*, vol. 50, nº 1, p. 53-63.

MAYS, N. et C. POPE (1995). « Rigour and qualitative research », *British Medical Journal*, vol. 311, nº 6997, p. 109-116.

MERIKANGAS, K.R., J.P. HE, M. BURSTEIN, S.A. SWANSON, S. AVENEVOLI, L. CUI *et al*. (2010). « Lifetime prevalence of mental disorders in US adolescents : Results from the national comorbidity study-adolescent supplement (NCS-A) », *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 49, nº 10, p. 980-989.

MONTAG, A., J.D. CLAPP, D. CALAC, J. GORMAN et C. CHAMBERS (2012). « A review of evidence-based approaches for reduction of alcohol consumption in native women who are pregnant or of reproductive age », *The American Journal of Drug and Alcohol Abuse*, vol. 38, nº 5, p. 436-443.

NAJAVITS, L.M., R.J. GALLOP et R.D. WEISS (2006). « Seeking safety therapy for adolescent girls with PTSD and substance use disorder : A randomized controlled trial », *The Journal of Behavioral Health Services & Research*, vol. 33, nº 4, p. 453-463.

PETTICREW, M. (2001). « Systematic reviews from astronomy to zoology : Myths and misconceptions », *British Medical Journal*, vol. 322, p. 98-101.

PETTICREW, M. et H. ROBERTS (2006). *Systematic Reviews in the Social Sciences. A Practical Guide*, Malden, Blackwell Publishing.

RIGGS, P.D., S.K. MIKULICH-GILBERTSON, R.D. DAVIES, M. LOHMAN, C. KLEIN et S.K. STOVER (2007). « A randomized controlled trial of fluoxetine and cognitive behavioral therapy in adolescents with major depression, behavior problems, and substance use disorders », *Archives of Pediatrics and Adolescent Medicine*, vol. 161, n° 11, p. 1026-1034.

SANTÉ CANADA (2002a). *Meilleures pratiques. Traitement et réadaptation des personnes aînées ayant des problèmes attribuables à la consommation d'alcool et d'autres drogues*, Ottawa, Gouvernement du Canada.

SANTÉ CANADA (2002b). *Meilleures pratiques. Troubles concomitants de santé mentale et d'alcoolisme et de toxicomanie*, Ottawa, Gouvernement du Canada.

SHEA, B.J., C. HAMEL, G.A. WELLS, L.M. BOUTER, E. KRISTJANSSON, J. GRIMSHAW *et al.* (2009). « AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews », *Journal of Clinical Epidemiology*, vol. 62, p. 1013-1020.

THOMAS, J. et A. HARDEN (2008). « Methods for the thematic synthesis of qualitative research in systematic reviews », *BMC Medical Research Methodology*, vol. 8, n° 45, <doi:10.1186/1471-2288-8-45>.

THURSTONE, C., P.D. RIGGS, S. SALOMONSEN-SAUTEL et S.K. MIKULICH-GILBERTSON (2010). « Randomized, controlled trial of atomoxetine for attention-deficit/hyperactivity disorder in adolescents with substance use disorder », *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 49, n° 6, p. 573-582.

TREMBLAY, J., K. BERTRAND, M. LANDRY et J. MÉNARD (2010). « Recension des écrits sur les meilleures pratiques en réadaptation auprès des adultes alcooliques ou toxicomanes », dans P. Desrosiers et J. Ménard (dir.), *Les services de réadaptation en toxicomanie auprès des adultes dans les centres de réadaptation en dépendance*, Montréal, Association des centres de réadaptation en dépendance du Québec, p. 11-87.

CHAPITRE 8

LA MÉTAANALYSE
Illustration pour déterminer si la toxicomanie aggrave les déficits cognitifs chez les personnes avec une schizophrénie

Stéphane Potvin

FORCES

- Elle a un caractère systématique et quantitatif, contrairement à une revue de littérature traditionnelle, de sorte qu'elle est mieux protégée contre les biais de confirmation.
- Elle confère généralement une excellente puissance statistique.
- Il s'agit d'une méthode simple et souple, qui s'applique à la grande majorité des types d'études réalisées en santé mentale.

LIMITES

- L'hétérogénéité des résultats peut compromettre la validité de la métaanalyse.
- Elle ne permet généralement pas d'effectuer des statistiques fines qui tiennent compte de l'ensemble des facteurs confondants pouvant influencer les résultats.
- L'évidence produite par une métaanalyse dépend étroitement de la bonne volonté des chercheurs de publier leurs résultats négatifs.

Sur le plan intellectuel, la publication en 1972 du livre *Effectiveness and Efficiency* d'Archie Cochrane constitue l'un des moments clés dans le développement de la médecine fondée sur les données probantes. Dans ce livre, Cochrane offre un plaidoyer en faveur de l'adoption, en milieu clinique, des interventions dont l'efficacité a été démontrée, préalablement, à l'aide d'essais randomisés et contrôlés (à double insu, avec placebo). C'est au cours des années 1990 que la mouvance de la médecine fondée sur les données probantes a connu son véritable envol, et qu'on a vu apparaître un nombre toujours croissant de publications relatant les résultats d'essais randomisés et contrôlés. Ces publications ont généré, en retour, le besoin de faire des synthèses des résultats de ces essais cliniques, par l'intermédiaire de métaanalyses. C'est ainsi qu'est né un réseau international visant à faire des synthèses de la connaissance, soit la Collaboration Cochrane, dont le nom a été choisi en guise de reconnaissance du caractère pionnier des idées formulées par Cochrane en 1972. La psychiatrie n'est pas demeurée en reste en ce qui concerne la médecine fondée sur les données probantes, si bien qu'elle est devenue l'un des domaines où il se réalise le plus grand nombre d'essais randomisés et contrôlés, de même qu'un des domaines où il se publie le plus grand nombre de métaanalyses. Dans le domaine de la santé mentale, la métaanalyse n'a pas seulement été utilisée en vue d'évaluer l'efficacité des interventions pharmacologiques et psychosociales ; elle a également été abondamment utilisée en vue de répondre à des questions épidémiologiques, cliniques, étiologiques et pathophysiologiques. Compte tenu de l'importance grandissante de la métaanalyse dans le domaine de la santé mentale, nous allons d'abord présenter, dans ce chapitre, les différentes étapes nécessaires à la réalisation d'une métaanalyse, tout en mettant en relief les forces et les faiblesses de cette méthode. Dans un second temps, nous allons présenter un exemple concret de métaanalyse réalisée dans le passé par notre groupe sur les effets cognitifs de la toxicomanie chez les personnes ayant la schizophrénie.

1. PRÉSENTATION DE LA MÉTAANALYSE

1.1. Définition

L'une des façons simples et relativement rapides de se familiariser avec un domaine d'études est de lire une revue de littérature publiée dans une revue scientifique avec comité de pairs. Bien que généralement instructif, cet exercice peut toutefois induire en erreur, puisque la revue de littérature reflète avant tout le point de vue *subjectif* des auteurs l'ayant rédigée. Divers biais peuvent compromettre les conclusions tirées d'une revue de littérature (voir

le chapitre 7 de cet ouvrage). Il peut d'abord y avoir un biais dans la *sélection* des études incluses dans la revue de littérature. La solution à ce problème est d'effectuer une revue de littérature dite *systématique*, où les auteurs décrivent la méthode par l'intermédiaire de laquelle ils ont effectué leur fouille des écrits. Lors de la rédaction d'une revue de littérature, plusieurs biais peuvent également se produire dans la *pondération* des évidences. Un auteur pourrait, par exemple, accorder beaucoup d'importance à une étude de faible qualité, mais dont les résultats sont conformes à l'hypothèse qu'il souhaite défendre au détriment d'une étude de grande qualité qui invalide au contraire cette hypothèse. La solution adoptée pour résoudre ce problème est d'effectuer une revue de littérature dite *quantitative*, faisant une synthèse statistique de toutes les évidences disponibles, par opposition à une revue qualitative. La métaanalyse correspond simplement à une revue de littérature qui est à la fois *systématique* et *quantitative* (<http://www.cochrane-net.org/openlearning/>). On pourrait penser que les biais ici invoqués font figure d'exception, mais, par exemple, le champ de la schizophrénie est au contraire rempli d'exemples de tels biais. Il suffit ici de penser aux multiples revues de littérature concluant à la supériorité nette des antipsychotiques de seconde génération par rapport aux antipsychotiques de première génération (Ichikawa et Meltzer, 1999 ; Meltzer, 1999). Les métaanalyses publiées sur cette question ont pourtant mis en évidence des résultats plus mitigés (Leucht *et al.*, 2009 ; Geddes *et al.*, 2000).

1.2. Forces de la métaanalyse

En plus d'être moins influencée par des biais de sélection et de pondération, la métaanalyse présente cette autre force d'avoir une plus grande puissance statistique que l'ensemble des articles publiés dans un domaine donné. L'un des problèmes fondamentaux de la recherche dans le domaine de la santé mentale est que la connaissance dans ce domaine se développe sur la base d'études dont les échantillons sont généralement petits. Par exemple, la schizophrénie, un trouble mental grave, s'accompagne de déficits cognitifs, de symptômes positifs (délires et hallucinations) et négatifs (manque de motivation, retrait social), d'une désorganisation conceptuelle et comportementale, d'une grande souffrance psychologique et de troubles de consommation (Nielssen *et al.*, 2012), qui rendent le recrutement difficile. Aussi, plusieurs des études publiées dans le domaine sont vulnérables à des biais d'échantillonnage, de sorte que leurs conclusions sont souvent difficilement généralisables à l'ensemble de la population d'intérêt. En regroupant toute la littérature disponible, la métaanalyse permet d'obtenir, dans bien des cas, une puissance statistique satisfaisante sur le plan scientifique. L'une des autres forces de la métaanalyse, moins discutée, mais pourtant fondamentale,

est qu'elle permet de déceler des sources d'hétérogénéité dans la littérature. Dans un domaine aussi complexe que celui de l'étude de la schizophrénie, il n'est pas rare que les méthodes utilisées et les résultats obtenus soient fort hétérogènes d'une étude à l'autre. Dans ces circonstances, les inférences faites à partir d'une métaanalyse perdent grandement de leur validité. Or, même dans ces circonstances, la métaanalyse peut être tout à fait utile, en révélant les sources de cette hétérogénéité, ce qui peut, en retour, avoir des conséquences méthodologiques importantes dans la conception des études qui seront effectuées dans le futur. Par exemple, notre équipe a effectué une métaanalyse à propos de l'effet de la toxicomanie sur les symptômes dépressifs des personnes aux prises avec la schizophrénie (Potvin, Sepehry et Stip, 2007). En effectuant la métaanalyse, nous nous sommes aperçus que le type d'échelle utilisée avait des répercussions directes sur les résultats, à savoir que les études utilisant l'inventaire de Beck (Beck *et al.*, 1961) ou l'échelle de Hamilton (1960) surestimaient l'effet de la toxicomanie sur les symptômes dépressifs, en comparaison de celles utilisant l'échelle de Calgary. Partant de ces résultats, nous avons donc émis la recommandation d'utiliser l'échelle de Calgary pour évaluer les symptômes dépressifs des personnes aux prises avec la schizophrénie et d'une toxicomanie cooccurrente. Enfin, l'une des forces de la méthode métaanalytique réside dans sa relative simplicité (comme nous allons le voir ci-dessous), qui fait en sorte que cette méthode d'analyse peut s'appliquer à une étonnante variété de types d'études, à la condition de respecter quelques principes statistiques de base.

1.3. Étapes à suivre pour réaliser une métaanalyse

Les étapes de la métaanalyse sont les suivantes (<http://www.cochrane-net.org/openlearning/>) : 1) la définition de la question de recherche ; 2) la fouille de la littérature à l'aide de logiciels de recherche de la littérature scientifique ; 3) l'inclusion ou l'exclusion des études en fonction de critères préétablis ; 4) l'extraction des données scientifiques ; 5) la réalisation du calcul de la taille de l'effet pour chacune des études incluses dans la métaanalyse ; 6) l'agrégation de la taille de l'effet de chacune des études incluses dans la métaanalyse ; 7) le calcul de l'hétérogénéité de la taille de l'effet de chacune des études incluses dans la métaanalyse ; 8) l'évaluation de la qualité de chacune des études incluses dans la métaanalyse (si possible) ; 9) le calcul de sous-analyses en fonction de variables modératrices continues ou catégorielles ; et 10) le calcul du biais de publication (tableau 8.1). La métaanalyse est effectuée à l'aide de logiciels, dont les plus utilisés sont, à notre connaissance, les logiciels Comprehensive Meta-Analysis-2, développé par Borenstein et Rothstein (1999), et RevMan (http://ims.cochrane.org/revman).

Tableau 8.1.
Les dix étapes de la métaanalyse

Les étapes	Conseils pratiques
1. La définition de la question de recherche	Quelle est la pertinence de la question ? A-t-elle déjà fait l'objet d'une métaanalyse ?
2. La fouille de la littérature	Ne pas se fier uniquement à PubMed.
3. L'établissement des critères d'inclusion et d'exclusion des études	Attention, ces choix peuvent directement influencer les résultats.
4. L'extraction des données scientifiques	Deux chercheurs/évaluateurs sont recommandés.
5. Le calcul de la taille de l'effet	$d = \frac{M_1 - M_2}{\delta}$ Attention, on ne peut effectuer un calcul de la taille de l'effet à partir de statistiques non paramétriques.
6. L'agrégation de la taille de l'effet de chacune des études incluses dans la métaanalyse	Il ne s'agit pas d'une simple moyenne des différentes tailles de l'effet. L'agrégation peut se faire à l'aide d'un modèle à effet fixe (qui tient compte de la taille d'échantillon des études) ou d'un modèle à effets aléatoires, qui tient compte de la taille des études et de la variabilité des résultats d'une étude à l'autre. En présence d'hétérogénéité, il importe d'utiliser le modèle à effets aléatoires.
7. Le calcul de l'hétérogénéité	Il s'agit d'un calcul de la variabilité des tailles de l'effet d'une étude à l'autre. En présence d'hétérogénéité, on peut se demander si les différentes études incluses dans la métaanalyse mesurent le même phénomène.
8. L'évaluation de la qualité des études	Utiliser des grilles validées d'évaluation de la qualité des études.
9. Les sous-analyses	Se font en fonction de variables continues (métarégression) ou catégorielles. Exemples de variables continues : âge, durée du traitement, etc. ; exemples de variables catégorielles : sexe, diagnostic, etc.
10. Le calcul du biais de publication	Le biais de publication se traduit par la relative absence de publication d'études dont l'échantillon est petit et dont les résultats sont négatifs.

Notes : ***d*** = taille de l'effet ; M_1 = moyenne du groupe 1 ; M_2 = moyenne du groupe 2 ; et δ = écart-type de la population formée par le groupe 1 et le groupe 2.

1.3.1. Définition de la question de recherche

Ici, il y a peu à dire sinon qu'il peut être pertinent de vérifier si notre idée a déjà fait l'objet d'une métaanalyse dans un logiciel de recherche tel que PubMed, et si un autre groupe est présentement en train de travailler sur la question, en consultant le site des revues systématiques Cochrane. Si la question de recherche concerne un traitement (pharmacologique ou non pharmacologique), il est recommandé d'enregistrer le protocole de recherche sur le site de Cochrane (<http://www.thecochranelibrary.com/view/0/index.html>).

1.3.2. Fouille de la littérature

Pour qu'une fouille de la littérature soit véritablement systématique, il importe de consulter plusieurs logiciels de recherche. L'une des erreurs fréquemment commises par les cliniciens et les chercheurs en médecine est de consulter exclusivement PubMed, alors que ce logiciel permet seulement de retracer 50 à 60% des études publiées dans le domaine de la médecine (Dickersin, Scherer et Lefebvre, 1994). Il est donc important de consulter d'autres logiciels, tels que PsycINFO, CINAHL, Embase, Web of Science, Cochrane, etc. Pour qu'une fouille soit de qualité, elle doit comprendre plusieurs mots clés, et il importe que ceux-ci ne soient pas trop spécifiques, mais plutôt généraux. De façon complémentaire, il est primordial, en cas de doute, de ne pas prendre de décision quant à l'inclusion ou à l'exclusion d'une étude sur la seule base de l'abrégé, mais de retracer l'article et de le lire. Si le chercheur s'intéresse, par exemple, aux fonctions exécutives dans la schizophrénie, et qu'il souhaite utiliser seulement les mots «fonctions exécutives» comme mots clés, il est fort probable que plusieurs études où les fonctions exécutives ont été évaluées dans la schizophrénie ne seront pas retenues, pour la simple et bonne raison que bon nombre d'études évaluent les fonctions exécutives dans la schizophrénie, sans en faire mention dans l'abrégé.

1.3.3. Inclusion ou exclusion des études

La sélection des articles doit être établie par consensus entre un minimum de deux chercheurs, selon des critères préétablis. Ces critères peuvent avoir une influence sur les résultats. Par exemple, une métaanalyse a été réalisée sur les inhibiteurs sélectifs de la recapture de la sérotonine (ISRS) dans la schizophrénie, et a montré que ces antidépresseurs n'améliorent pas les symptômes négatifs de la schizophrénie (Sepehry *et al.*, 2007). Une autre métaanalyse a montré, au contraire, que les antidépresseurs peuvent améliorer de façon significative les symptômes négatifs de la schizophrénie

(Rummel, Kissling et Leucht, 2005). Pourquoi une telle différence dans les résultats ? Il est possible que cette différence soit due au fait que dans la première métaanalyse (Sepehry *et al.*, 2007), seuls les ISRS ont été considérés, alors que la métaanalyse de Rummel *et al.* (2005) a inclus tous les types d'antidépresseurs.

1.3.4. Extraction des données scientifiques

L'extraction des données doit se faire par un minimum de deux chercheurs, afin de s'assurer qu'il n'y a aucune erreur d'entrée de données.

1.3.5. Calcul de la taille de l'effet

Le calcul de la taille de l'effet dépend d'une formule simple, soit :

$$d = \frac{M_1 - M_2}{\delta}$$

où

d = taille de l'effet ;

M_1 = moyenne du groupe 1 ;

M_2 = moyenne du groupe 2 ;

δ = écart-type de la population formée par le groupe 1 et le groupe 2.

La taille de l'effet correspond donc à une différence de moyenne standardisée (Cohen, 1988). En ce sens, la taille de l'effet est très similaire au score *z*, à cette différence près que le score *z* s'intéresse à la déviation d'un individu par rapport à son groupe d'appartenance, alors que la taille de l'effet s'intéresse à la déviation d'un groupe par rapport à un autre. Dans les deux cas, l'écart-type correspond en quelque sorte à l'« unité de mesure ». Par convention, on estime qu'une taille de l'effet de 0,2 est faible ; de 0,5, modérée ; et de 0,8, forte (Cohen, 1988).

La taille de l'effet peut être calculée à partir de la moyenne et de l'écart-type des groupes qui font l'objet de la comparaison. Elle peut aussi être calculée à partir d'un test *t* à groupes indépendants ou à groupes appariés ; de coefficients de corrélation de Pearson ou de coefficients de régression linéaire ; ou encore d'analyses de variance (ANOVA) à groupes indépendants, à mesures répétées ou d'ANOVA mixtes. Des métaanalyses peuvent également être réalisées lorsque la variable dépendante d'intérêt est dichotomique. Dans ce cas, le calcul de la taille de l'effet peut être dérivé d'un ratio de cote (*odds ratio*), d'un risque relatif ou d'un chi-carré. Comme

on le voit, la métaanalyse est très souple sur le plan statistique, mais il importe d'avoir à l'esprit qu'elle ne peut s'appliquer dans le cas d'analyses statistiques non paramétriques, telles que le test de Mann-Whitney (groupes indépendants), le test de Wilcoxon (mesures répétées) ou les coefficients de corrélation de Spearman.

1.3.6. Agrégation des tailles de l'effet

L'agrégation des tailles de l'effet ne correspond pas à une simple moyenne des tailles de l'effet de chacune des études incluses dans la métaanalyse. L'agrégation des tailles de l'effet peut se faire selon deux modèles principaux, soit le modèle à effets fixes et le modèle à effets aléatoires (*random effects*) (DerSimonian et Laird, 1986). Dans le modèle à effets fixes, l'agrégation tient compte de la précision de chacune des études, laquelle dépend étroitement, en retour, de la taille de l'échantillon recruté. Ainsi, dans le cadre de la métaanalyse, une étude comprenant 1000 participants aura un poids relatif nettement plus lourd dans le calcul de la taille de l'effet composé (*composite effect size*) qu'une étude comprenant seulement 50 participants. Dans le modèle à effets aléatoires, la taille de l'effet composé est pondérée en fonction non seulement de la précision de chaque étude comprise dans la métaanalyse, mais également de la déviation de chaque étude par rapport à l'effet composé. Ce dernier point nous amène à l'épineuse question de l'hétérogénéité.

1.3.7. Calcul de l'hétérogénéité des tailles de l'effet

Après avoir calculé l'effet composé, il faut déterminer si les résultats de la métaanalyse sont homogènes ou hétérogènes d'une étude à l'autre. L'une des façons standards de calculer l'hétérogénéité est le Q de Cochrane (*Cochrane's Q*), lequel évalue la variation typique des tailles de l'effet des études individuelles par rapport à l'effet composé (Higgins *et al.*, 2003). Par convention, lorsque la valeur p (niveau de signification) produite en calculant le Q de Cochrane est inférieure à 0,1, on estime alors qu'on est en présence d'hétérogénéité. Plus la valeur p est petite, et plus on a de raisons de penser que la métaanalyse regroupe des études qui ne sont pas comparables, et par conséquent, que la métaanalyse n'est pas valide.

En présence d'hétérogénéité, trois solutions peuvent être adoptées. La première est de reconsidérer la pertinence de la métaanalyse, la question de recherche ou les critères d'inclusion et d'exclusion qui ont été utilisés. Cela dit, la grande majorité des chercheurs n'applique pas cette solution. La deuxième solution est d'adopter un modèle à effets aléatoires afin de

calculer la taille de l'effet composé. Il importe toutefois de préciser que cette dernière solution ne règle pas le problème de l'hétérogénéité sur le plan *méthodologique*, et qu'elle le pondère seulement par une procédure *statistique*. Enfin, en présence d'hétérogénéité, il est possible d'effectuer des sous-analyses, afin de tenter de trouver les sources de cette hétérogénéité, qu'il s'agisse de variables catégorielles ou continues.

1.3.8. Sous-analyses

Les sous-analyses impliquant des variables catégorielles consistent simplement à regrouper des études incluses dans la métaanalyse (ou de sous-populations à l'intérieur de ces études) en fonction de caractéristiques discontinues des populations étudiées. Par exemple, une métaanalyse a été réalisée sur la perception de la douleur dans la schizophrénie, telle que mesurée en contexte expérimental, laquelle a démontré que les personnes aux prises avec la schizophrénie ont une réponse diminuée à la douleur. Afin de savoir si cette hypoalgésie est attribuable à l'effet des antipsychotiques, une sous-analyse a été effectuée, où les études ont été regroupées en fonction de la prise d'antipsychotiques au moment de l'expérimentation, ce qui a permis de montrer que les personnes aux prises avec la schizophrénie ne perçoivent pas normalement la douleur, qu'elles soient traitées ou non avec des antipsychotiques au moment de l'expérimentation.

Les variables continues peuvent être également très utiles lorsque l'on réalise une métaanalyse. Lorsque la situation est pertinente et le permet, on peut effectuer une analyse de métarégression. Il s'agit d'une analyse de régression pondérée en fonction de la précision des études (lorsque la métarégression est effectuée à l'aide d'un modèle à effets fixes) et de la variabilité interétudes (lorsque la métarégression est effectuée à l'aide d'un modèle à effets aléatoires) (<http://www.cochrane-net.org/openlearning/>). Par exemple, il est bien connu que les antipsychotiques de seconde génération (ASG) produisent généralement moins de symptômes neurologiques moteurs (p. ex., le parkinsonisme) que les antipsychotiques de première génération (APG). Des auteurs ont toutefois remarqué que le dosage de l'halopéridol, un APG, est parfois trop élevé dans les essais randomisés comparant les ASG et les APG. À l'aide d'analyses de métarégression tenant compte du dosage de l'halopéridol (variable continue), certaines métaanalyses ont montré que les ASG produisent effectivement moins d'effets secondaires neurologiques que les APG, mais que cette différence est moins prononcée lorsque l'on considère le dosage de l'halopéridol (Geddes *et al.*, 2000).

1.3.9. Évaluation de la qualité des études incluses dans la métaanalyse

L'une des variables continues qu'il importe de mesurer, surtout dans les métaanalyses portant sur des essais cliniques, est la qualité des études individuelles incluses dans la métaanalyse. Une fois la qualité évaluée, on peut effectuer des analyses de métarégression afin de savoir si l'effet composé obtenu dépend ou non de la qualité des études retenues. L'évaluation de la qualité des études doit toutefois se faire en fonction de grilles d'évaluation validées, ce qui n'est pas toujours possible. Pour les métaanalyses en imagerie cérébrale ou les métaanalyses portant sur des biomarqueurs, de telles grilles n'existent pas, du moins à notre connaissance. En revanche, de telles grilles existent quand il s'agit d'évaluer des essais contrôlés (voir le chapitre 10 de cet ouvrage). Parmi les critères d'évaluation, il y a entre autres la présence ou l'absence d'une procédure de randomisation, le fait que l'étude soit à simple insu, à double insu ou ouverte, la réalisation d'un calcul de puissance statistique, ou encore l'inclusion des participants qui abandonnent l'étude avant la fin du traitement dans les analyses statistiques (<http://www.cochrane-net.org/openlearning/>).

1.3.10. Calcul du biais de publication

La validité de toute métaanalyse dépend, en partie, de la bonne volonté des chercheurs de publier *tous* leurs résultats, incluant les résultats de recherche défavorables à leurs propres hypothèses. Or, la nature humaine étant ce qu'elle est, il arrive trop souvent que les chercheurs décident de ne pas rapporter les résultats négatifs ou contraires à leurs attentes. Similairement, bon nombre de revues scientifiques ont tendance à favoriser les résultats positifs et à ne pas publier les résultats négatifs. Afin de pallier ce problème fondamental, il est possible d'estimer le nombre d'études qui seraient nécessaires afin d'invalider les résultats de la métaanalyse (en anglais, on parle du *fail-safe*). Ce type de calcul est toutefois peu fiable.

Comme alternative, il est possible d'estimer, à l'aide de divers tests tels que le test d'Egger (Leandro, 2005), la présence ou l'absence d'un biais de publication dans la littérature disponible. Bien que ces tests ne soient pas infaillibles, ils reposent sur une logique facile à comprendre. En règle générale, les études comprenant de larges échantillons sont faciles à publier, que les résultats de l'étude soient positifs ou non. En effet, on voit difficilement comment un éditeur d'une revue scientifique pourrait refuser de publier un essai randomisé incluant 5000 participants au seul motif que les résultats de l'étude sont négatifs. Les études rapportant des résultats positifs, mais comprenant des petits échantillons se publient généralement bien aussi (quoique dans des revues dont le facteur d'impact est moins élevé), car elles

peuvent générer de nouvelles hypothèses ou de nouvelles avenues thérapeutiques dont on ne voudrait pas faire abstraction. Ce sont toutefois des études typiquement associées à des tailles de l'effet élevées, car, sans cela, ces études n'auraient pas eu la puissance statistique pour générer des résultats significatifs. Là où le bât blesse, c'est dans le cas des études *négatives* dont les résultats ont été obtenus sur la base de petits échantillons, mais qui ont été réalisées avec rigueur. Même lorsque les chercheurs sont motivés à les rapporter (ce qui n'est pas la norme), ces résultats sont difficiles à publier, pour la simple et bonne raison qu'il est malaisé de conclure à l'inefficacité d'une intervention donnée si l'essai clinique ne comprenait pas, au point de départ, un échantillon de participants suffisamment grand pour détecter un effet bénéfique de l'intervention. Ainsi, on devrait s'attendre, en présence d'un biais de publication, à ce que plusieurs études négatives avec des petits échantillons n'aient pas été rapportées, et que seules les grandes études (c.-à-d. une taille d'échantillon élevée) et les petites études positives l'aient été. Par conséquent, en présence d'un biais de publication, on devrait observer une relation *négative* entre la taille de l'effet des études et la taille de leurs échantillons : plus l'étude est petite et plus la taille de l'effet est grande (Egger *et al.*, 1997). Inversement, lorsqu'il n'y a pas de biais de publication (et que les petites études négatives sont publiées), on devrait s'attendre à ne pas observer de relation entre la taille de l'étude et la taille de son effet. Sans surprise, les tests qui permettent de mesurer le biais de publication sont essentiellement des analyses de régression, adaptée au contexte métaanalytique, qui examinent la relation entre la taille de l'effet des études individuelles et la précision de ces études (laquelle dépend étroitement de la taille de l'échantillon).

En définitive, l'épineux problème du biais de publication ne pourra être résolu sans un changement de culture de la publication. Dans le domaine de la psychopharmacologie, l'instauration d'un registre d'essais cliniques (<http://www.clinicaltrials.gov>) constitue un pas dans cette direction. En effet, les chercheurs dans ce domaine ont l'obligation d'enregistrer leurs essais pharmacologiques avant de les entamer, et afin de s'en assurer, un nombre grandissant de revues scientifiques exigent le numéro d'enregistrement de l'étude, sans quoi l'article qui en dérive ne peut pas être publié. L'existence de ce registre fait en sorte qu'il est plus facile pour les experts en métaanalyse de repérer des études dont les résultats n'ont pas été publiés, et ils peuvent dès lors contacter les auteurs concernés et leur demander leurs résultats. À notre connaissance, un tel registre n'existe pas pour les interventions psychosociales, mais un nombre grandissant de chercheurs dans ce domaine prennent maintenant l'habitude d'enregistrer leurs études sur le site <http://www.clinicaltrials.gov>.

1.4. Limites de la métaanalyse

Les deux limites fondamentales de la métaanalyse sont celles de l'hétérogénéité des résultats et de la présence de biais de publication. La métaanalyse présente toutefois d'autres limites, moins discutées, mais qui peuvent néanmoins influencer de façon importante les résultats. L'un de ces problèmes est celui de la sélection de statistiques ajustées ou non ajustées quand vient le temps de calculer la taille de l'effet composé. En psychiatrie, il arrive fréquemment que les résultats soient influencés par des variables confondantes qui doivent être contrôlées, ne serait-ce que statistiquement. Ce qui devient problématique dans un contexte métaanalytique, c'est lorsque les variables confondantes considérées ne sont pas les mêmes d'une étude à l'autre. Les chercheurs doivent alors choisir entre le fait d'utiliser des statistiques ajustées ou non ajustées afin de calculer la taille de l'effet composé. Les uns vont rejeter les statistiques non ajustées, car elles surestiment généralement la taille des effets, alors que les autres vont rejeter les statistiques ajustées afin d'éviter de faire l'agrégation d'études qui ne sont pas comparables dans leurs contrôles statistiques. Face à ce dilemme, il n'y a pas de réponse toute faite, mais chose sûre, ce choix peut avoir une influence directe sur les résultats. Il a été montré, par exemple, que le fait de fumer du cannabis augmente de 2,9 fois le risque de développer des symptômes psychotiques si on ne considère pas l'influence de diverses variables confondantes (l'âge de début de consommation, les traits de personnalité prémorbide, etc.), mais que ce risque est augmenté de seulement 1,4 si on les considère (Moore *et al.*, 2007 ; Semple *et al.*, 2005).

Un autre problème de la méthode métaanalytique parfois occulté réside dans le fait que la valeur de la métaanalyse est directement tributaire du choix d'analyses statistiques fait par les différents auteurs, alors qu'elles ne sont pas forcément optimales. Ce problème refait régulièrement surface dans le cas des métaanalyses portant sur des études d'intervention. Dans les essais cliniques, les analyses statistiques devraient normalement être mixtes, c'est-à-dire qu'elles devraient tenir compte de l'effet du passage du temps, de l'effet du traitement et de leur interaction. Or il arrive régulièrement dans les essais cliniques comprenant de petits échantillons que les auteurs adoptent des analyses statistiques plus rudimentaires, qui augmentent la probabilité de détecter des effets thérapeutiques. Dans les essais cliniques comprenant de petits échantillons, les auteurs appliquent régulièrement des tests t ou des analyses de variance sur les données des participants en fin de traitement, sans tenir compte des données des participants en début de traitement. Mine de rien, ce choix statistique peut biaiser les résultats à la hausse, comme nous avons eu l'occasion de le montrer dans la métaanalyse sur les ISRS dans la schizophrénie évoquée à la section 1.3.3 (Sepehry *et al.*, 2007).

1.5. Multiples applications de la métaanalyse

En dépit de ces limites, la méthode métaanalytique est de plus en plus utilisée en psychiatrie, en raison de sa simplicité et de sa souplesse. Par exemple, en date du 1er juillet 2013, on dénombre plus de 1300 métaanalyses qui ont été publiées dans le seul domaine de la schizophrénie. Les applications de la métaanalyse dans le domaine de la schizophrénie sont multiples et variées, et il s'agit d'une méthode qui a été utilisée afin de répondre à des questions épidémiologiques, cliniques, pathophysiologiques et thérapeutiques (Leon et Diaz, 2005; Fusar-Poli *et al.*, 2012; Howes *et al.*, 2012; Marsman *et al.*, 2013; Moore *et al.*, 2007; Mössler *et al.*, 2011; Roder, Mueller et Schmidt, 2011; Wykes *et al.*, 2011).

2. EXEMPLE D'APPLICATION DE LA MÉTAANALYSE DANS LE DOMAINE DE LA SCHIZOPHRÉNIE

2.1. Question de recherche

La schizophrénie est un trouble mental grave qui s'accompagne de déficits cognitifs notables. De fait, entre 70 et 75% des personnes atteintes de la schizophrénie ont des déficits cognitifs, et ces déficits se manifestent dans des domaines variés comme l'attention, les fonctions exécutives, la mémoire épisodique, la mémoire de travail, la vitesse de traitement de l'information et la cognition sociale (Green *et al.*, 2004; Savla *et al.*, 2013). Les déficits cognitifs de la schizophrénie prédisent mieux le fonctionnement social et occupationnel de ces personnes que les symptômes positifs et négatifs. La toxicomanie est l'un des facteurs qui devrait normalement aggraver les déficits cognitifs associés à la schizophrénie (Fett *et al.*, 2011). À vie, près de 50% des personnes atteintes de la schizophrénie vont présenter une toxicomanie. Dans l'ordre décroissant et si on exclut le tabagisme, ces personnes consomment du cannabis, de l'alcool ou des psychostimulants. Or il est bien connu que la consommation chronique d'alcool, de cannabis ou de psychostimulants peut engendrer des déficits cognitifs importants (Crean, Crane et Mason, 2011; Potvin *et al.*, soumis; Stavro, Pelletier et Potvin, 2013). En dépit de cela, les études neuropsychologiques qui se sont intéressées aux effets de la toxicomanie sur la cognition des personnes atteintes de la schizophrénie n'ont pas produit les résultats escomptés. En effet, alors que certaines études ont bel et bien montré que la toxicomanie aggrave la performance cognitive des personnes atteintes de la schizophrénie, d'autres études ont montré une absence d'effets, et d'autres études ont montré, enfin, que la toxicomanie est associée à une *meilleure* performance cognitive

(Potvin, Stavro et Pelletier, 2012). En présence de ces évidences contradictoires, nous avons décidé d'effectuer une métaanalyse de la littérature disponible afin de déterminer si la toxicomanie aggrave les déficits cognitifs de la schizophrénie, et afin d'identifier les facteurs pouvant expliquer le caractère contradictoire des résultats (Potvin *et al.*, 2008a). Pour toutes les analyses, nous avons utilisé le logiciel Comprehensive Meta-Analysis-2, développé par Borenstein et Rothstein (1999).

2.2. Fouille de la littérature

Une fouille des bases informatisées de la littérature scientifique PubMed et PsycINFO a été réalisée, à l'aide des mots clés suivants (en anglais): «schizophrénie» et «alcool» ou «amphétamine» ou «cannabis» ou hallucinogène» ou «héroïne» ou «marijuana» ou «phencyclidine» et «cognition». Cette fouille a été complétée en consultant la liste des références bibliographiques citées par chacune des études repérées.

2.3. Critères d'inclusion et d'exclusion

Les critères d'inclusion étaient les suivants: 1) toute étude comprenant des personnes aux prises avec un trouble de la lignée de la schizophrénie; 2) des personnes atteintes de la schizophrénie avec et sans toxicomanie cooccurrente (abus ou dépendance); 3) des personnes ayant un problème de consommation d'alcool, d'amphétamines, de cannabis, de cocaïne, d'hallucinogènes, d'héroïne ou de phencyclidine; 4) toute étude mesurant la cognition à l'aide de tests neuropsychologiques validés; et 5) toute étude publiée avant le 31 août 2006. Les critères d'exclusion étaient les suivants: 1) toute étude comprenant des personnes souffrant d'un trouble de l'humeur uni- ou bipolaire; et 2) des personnes atteintes de la schizophrénie consommant exclusivement la cigarette ou des benzodiazépines.

2.4. Problème de la classification des tests cognitifs

Avant d'effectuer les calculs métaanalytiques à proprement parler, nous avons rencontré un problème fondamental auquel se bute quiconque tente d'effectuer une métaanalyse neuropsychologique, soit celui de la classification des tests cognitifs. Ce problème se pose, d'une part, parce que les tests cognitifs utilisés pour mesurer un même domaine cognitif varient beaucoup d'une étude à l'autre et, d'autre part, parce qu'il n'y a pas de consensus, dans le cas de plusieurs tests cognitifs, quant au principal domaine cognitif

mesuré par ces tests. Un exemple classique est celui du test d'interférence de Stroop (MacLeod, 1991). Selon certains auteurs, il s'agit avant tout d'une tâche d'attention sélective; pour d'autres, il s'agit essentiellement d'une tâche mesurant la capacité d'inhiber des réponses surapprises; alors que d'autres, enfin, estiment plutôt qu'il s'agit d'une tâche mesurant principalement la vitesse de traitement de l'information. Afin de résoudre ce problème, nous nous sommes fiés aux conclusions du groupe consensuel MATRICS (Measurement and Treatment Research to Improve Cognition in Schizophrenia) (Green *et al.*, 2004). Mis sur pied par le National Institute of Health, le groupe d'experts MATRICS s'est donné comme objectif de déterminer les principaux domaines cognitifs atteints dans la schizophrénie et de créer une batterie cognitive permettant de mesurer rapidement l'ensemble de ces domaines. Sur la base d'analyses factorielles effectuées dans le cadre de 13 études neuropsychologiques menées chez des personnes atteintes de la schizophrénie, ce groupe d'experts est arrivé à la conclusion que sept principaux domaines cognitifs sont atteints dans la schizophrénie, soit l'attention, les fonctions exécutives, la mémoire verbale, la mémoire visuelle, la mémoire de travail, la vitesse de traitement de l'information et la cognition sociale. Fait intéressant, les conclusions tirées par ce groupe d'expert permettent de classifier, dans le domaine cognitif approprié, une grande variété de tests cognitifs utilisés dans la schizophrénie. Aussi, nous nous sommes fiés aux conclusions de ce groupe afin de classifier les différents tests utilisés dans les études que nous avons incluses dans la métaanalyse. Dans le cas des tests auxquels le groupe MATRICS ne fait pas référence, nous avons procédé à une fidélité interjuges entre deux membres de l'équipe de recherche afin de classer le ou les tests en question en fonction de leur similitude conceptuelle avec des tests auxquels le groupe MATRICS fait référence.

2.5. Calcul métaanalytique global

Un total de 22 articles comprenant 1870 personnes atteintes de la schizophrénie (avec et sans toxicomanie cooccurrente) ont été repérés. Dans un premier temps, nous avons effectué un calcul de la taille de l'échantillon pour la performance *globale* des deux groupes de personnes dans chaque étude, *indépendamment* des domaines cognitifs mesurés. Ici, le calcul de la taille de l'effet correspond à la différence standardisée de la performance cognitive moyenne des participants avec et sans toxicomanie cooccurrente. Pour chaque étude, nous avons calculé un «index cognitif global», en faisant tout simplement la moyenne des tailles de l'effet de chaque domaine testé dans chaque étude individuelle, puis nous avons fait l'agrégation des résultats de l'ensemble des études. En présence d'hétérogénéité dans nos résultats, nous avons effectué l'ensemble des calculs métaanalytiques à l'aide d'un modèle

à effets aléatoires, lequel tient compte de la variance intra- et interétudes. Pour cette analyse, nos résultats ont montré une absence de différence dans la performance cognitive entre les personnes atteintes de la schizophrénie avec et sans toxicomanie cooccurrente : *g* de Hedges = 0,06 ; limite inférieure = –0,079 ; limite supérieure : 0,199 ; *p* = 0,394. Il est à noter que le *g* de Hedges est l'équivalent du *d* de Cohen, mais qu'il est moins influencé par les études comprenant des petits échantillons. Nous avons décidé d'utiliser le *g* de Hedges plutôt que le *d* de Cohen, parce que notre métaanalyse comprenait un certain nombre d'études incluant moins de 35 participants au total. Il est aussi à noter qu'un *g* de Hedges positif signifie que le groupe schizophrénie et toxicomanie a une meilleure performance cognitive que le groupe schizophrénie sans toxicomanie.

2.6. Sous-analyses

2.6.1. Sous-analyses catégorielles

L'objectif fondamental de la métaanalyse étant moins de déterminer si la toxicomanie est associée à plus ou moins de déficits cognitifs dans la schizophrénie que de déterminer les facteurs expliquant l'hétérogénéité des résultats d'une étude à l'autre, nous avons effectué des sous-analyses en utilisant des variables catégorielles. Dans un premier temps, nous avons analysé les différences entre les participants avec et sans toxicomanie cooccurrente sur des tests cognitifs précis (p. ex., Trail-Making Test, Wisconsin Card Sorting Test, Controlled Oral Word Association Test, etc.) ayant été utilisés dans un nombre d'études suffisamment grand. Pour ce faire, nous avons calculé une taille de l'effet pour chaque étude ayant utilisé la tâche d'intérêt, puis nous avons fait l'agrégation des résultats de l'ensemble des études à l'aide d'un modèle à effets aléatoires, en raison de la présence d'hétérogénéité d'une étude à l'autre. Les résultats ont montré que les personnes avec un double diagnostic de schizophrénie et de toxicomanie ont moins de déficits sur le Trail-Making Test (version A : *g* de Hedges = 0,506 ; *p* = 0,0001 ; version B : *g* de Hedges = 0,524 ; *p* = 0,0001), soit une tâche qui mesure principalement la vitesse de traitement de l'information dans la schizophrénie, et que cette différence correspond à une taille de l'effet modérée. Pour les autres tests utilisés, aucune différence significative entre les groupes n'a été détectée. Dans le même ordre d'idées, nous avons vérifié si le domaine cognitif étudié est l'un des facteurs qui expliquent l'hétérogénéité des résultats publiés à ce jour. Dans le cas de la vitesse de traitement de l'information, nous avons eu l'occasion de faire l'agrégation des résultats de 16 études (*n* total = 1245) mesurant ce domaine cognitif (peu importe le test utilisé), et les résultats

ont montré, ici encore, que les personnes ayant un double diagnostic ont légèrement moins de ralentissement psychomoteur (*g* de Hedges = 0,211 ; p = 0,037). Dans le cas des autres domaines cognitifs, nous n'avons pas détecté de différences significatives entre les groupes.

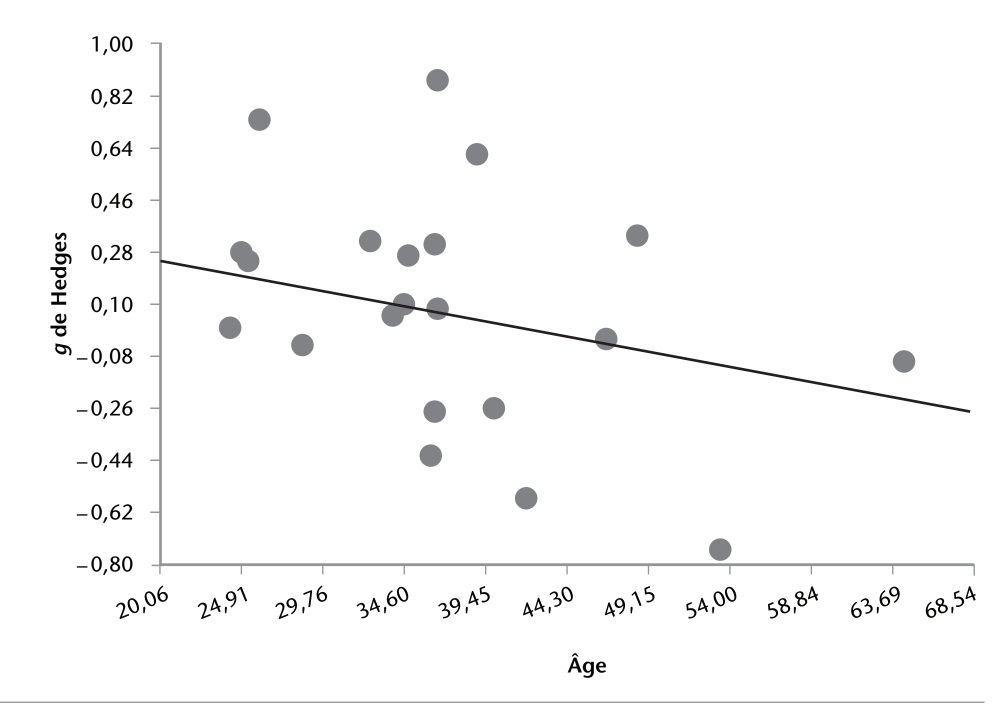
Dans un deuxième temps, nous avons voulu savoir si le type de substances consommées peut expliquer une partie de l'hétérogénéité des résultats. Nous avons donc effectué des sous-analyses où nous avons regroupé les études incluses dans la métaanalyse en fonction de la substance consommée dans chacune des études, soit l'alcool, le cannabis ou la cocaïne. Nous avons regroupé dans la catégorie « consommation mixte » les études où la majorité des personnes avait une polyconsommation. En effectuant ces sous-analyses, nous avons pu montrer, en utilisant l'*index cognitif global*, que la consommation est associée à une relative préservation cognitive dans la schizophrénie (*g* de Hedges = 0,571 ; p = 0,001). Par ailleurs, nous avons aussi pu montrer que l'alcool semble aggraver les déficits de mémoire de travail dans la schizophrénie (*g* de Hedges = –0,415 ; p = 0,034) et que la consommation de cannabis va de pair avec une relative préservation cognitive sur les plans des fonctions exécutives (*g* de Hedges g = 0,789 ; p = 0,0001) et de la mémoire visuelle (*g* de Hedges = 0,446 ; p = 0,011). Ces analyses ayant été réalisées sur un petit nombre d'études ($n \leq 3$), ces résultats doivent assurément être considérés comme étant préliminaires. Pris ensemble, ces résultats laissent penser, de façon préliminaire, que le type de substances semble effectivement influencer les résultats des études mesurant la performance cognitive chez les personnes atteintes de schizophrénie.

2.6.2. Analyse de métarégression

En plus des sous-analyses catégorielles, nous avons effectué des sous-analyses utilisant des variables continues. Notre question de recherche consistait à savoir si la durée de la toxicomanie pouvait exercer un effet sur la performance cognitive des personnes atteintes de schizophrénie. L'idée était de voir si la toxicomanie serait associée à une détérioration cognitive dans la schizophrénie, principalement en présence d'une consommation chronique de substances psychoactives. Malheureusement, nous avons constaté que la grande majorité des études incluses dans la métaanalyse n'ont pas mesuré la durée de la toxicomanie chez les personnes atteintes de schizophrénie, ce qui a fait en sorte que nous avons dû nous rabattre sur une analyse moins précise, portant sur la relation potentielle entre l'âge des personnes atteintes d'un trouble comorbide et la taille de l'effet de chaque étude. Nous avons donc extrait de chaque étude l'âge moyen du groupe de personnes présentant la schizophrénie et une toxicomanie cooccurrente, et nous avons effectué des calculs de métarégression. Dans ces calculs, l'âge moyen des

personnes ayant les deux conditions a été utilisé comme variable prédictive de la taille de l'effet obtenue dans chaque étude. Plus précisément, la relation entre l'âge moyen des personnes ayant un trouble comorbide et la taille de l'effet a été examinée, en commençant par la différence dans la performance cognitive globale (index cognitif global) entre les patients avec et sans toxicomanie cooccurrente. Des analyses de métarégression ont également été effectuées dans le cas des domaines cognitifs ayant été mesurés dans 10 études ou plus. Les résultats ont montré une relation linéaire négative entre la taille de l'effet et l'âge moyen des personnes ayant un trouble comorbide, tant dans le cas de l'index cognitif global (21 études; $\beta = -0{,}011$; $p = 0{,}040$) que dans le cas de la vitesse de traitement de l'information (15 études; $\beta = -0{,}029$; $p = 0{,}001$) et de la mémoire de travail (10 études; $\beta = -0{,}034$; $p = 0{,}0001$). Comme indiqué dans la figure 8.1, qui porte sur l'index cognitif global, les analyses de métarégression ont montré

Figure 8.1.
Âge des personnes atteintes de schizophrénie comme variable prédictive de la taille de l'effet (index cognitif composé)



Source : Potvin *et al.*, 2008a.

que la performance cognitive des participants ayant un trouble comorbide est supérieure à celle des participants non toxicomanes avant l'âge de 35 ans environ. Chez les participants âgés de plus de 35 ans, on observe toutefois un renversement de tendance, à savoir qu'à partir de cet âge, les participants ayant un trouble comorbide ont une performance cognitive inférieure à celle des participants non toxicomanes. Ces résultats indiquent soit que la toxicomanie produit ses effets délétères sur la cognition dans la schizophrénie à partir de l'âge de 30 ans (environ) ou encore que la toxicomanie produit ses effets délétères sur la cognition seulement après une consommation chronique de substances psychoactives. Cette seconde interprétation tient pour acquis qu'il existe une relation étroite entre l'âge des personnes atteintes de la schizophrénie et la durée de leur consommation, ce qui semble probable, mais qui demeure à démontrer.

2.7. Biais de publication

Comme il ne s'agissait pas d'une métaanalyse portant sur des études d'intervention, nous n'avons pas effectué de calcul de biais de publication. Mentionnons seulement ici que le simple fait que les auteurs aient publié des résultats contradictoires et des résultats négatifs indique, en soi, que la probabilité d'un biais de publication est plutôt faible.

CONCLUSION

Dans le cadre de la médecine fondée sur les données probantes, on considère généralement que la métaanalyse est la méthode dont les résultats offrent le niveau le plus élevé d'évidence empirique. Il s'agit d'une méthode simple et souple qui s'applique à la quasi-totalité des types de recherche effectués dans le domaine de la santé mentale. Cette méthode confère généralement une excellente puissance statistique. De plus, en raison de son caractère systématique et quantitatif, cette méthode est généralement mieux protégée contre les biais de confirmation que ne l'est une revue de littérature traditionnelle. La validité de la métaanalyse est toutefois fragilisée par la présence d'hétérogénéité dans les résultats, ce qui constitue la norme dans le complexe domaine de la santé mentale. La qualité des évidences produites par la métaanalyse dépend, par ailleurs, de la bonne volonté des scientifiques de publier leurs résultats négatifs, ce qui ne fait pas présentement partie des mœurs et coutumes en recherche en psychiatrie. Finalement, en raison de son caractère macroscopique, la métaanalyse ne permet pas de faire des analyses statistiques fines, de sorte qu'une métaanalyse regroupant 20 études comprenant un total de 400 participants ne

remplacera jamais une étude de grande envergure incluant 400 participants où toutes les variables potentiellement confondantes sont adéquatement contrôlées. Dans le futur, il faudra trouver des façons novatrices de changer la culture actuelle de la publication, détecter les sources qui contribuent à l'hétérogénéité des résultats de plusieurs métaanalyses, et veiller à ce que les résultats de la médecine fondée sur les données probantes ne restent pas sur des tablettes et servent au contraire à offrir des meilleurs traitements aux personnes vivant avec divers troubles mentaux.

RÉFÉRENCES

BECK, A.T., C.H. WARD, M. MENDELSON, J. MOCK et J. ERBAUGH (1961). « An inventory for measuring depression », *Archives of General Psychiatry*, vol. 4, p. 561-571.

BORENSTEIN, M. et H. ROTHSTEIN (1999). *Comprehensive Meta-Analysis : A Computer Program for Research Synthesis*, Englewood, Biostat.

COCHRANE, A. (1972). *Effectiveness and Efficiency : Random Reflections on Health Services*, Londres, Nuffield Provincial Hospitals Trust.

COHEN, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2e éd., Hillsdale, Erlbaum.

CREAN, R.D., N.A. CRANE et B.J. MASON (2011). « An evidence based review of acute and long-term effects of cannabis use on executive cognitive functions », *Journal of Addiction Medicine*, vol. 5, no 1, p. 1-8.

DERSIMONIAN, R. et N. LAIRD (1986). « Meta-analysis in clinical trials », *Controlled Clinical Trials*, vol. 7, p. 177-188.

DICKERSIN, K., R. SCHERER et C. LEFEBVRE (1994). « Identifying relevant studies for systematic reviews », *British Medical Journal*, vol. 309, no 6964, p. 1286-1291.

EGGER, M., G. DAVEY SMITH, M. SCHNEIDER et C. MINDER (1997). « Bias in meta-analysis detected by a simple, graphical test », *British Medical Journal*, vol. 315, no 7109, p. 629-634.

FETT, A.K., W. VIECHTBAUER, M.D. DOMINGUEZ, D.L. PENN, J. VAN OS et L. KRABBENDAM (2011). « The relationship between neurocognition and social cognition with functional outcomes in schizophrenia : A meta-analysis », *Neuroscience and Biobehavioral Reviews*, vol. 35, no 3, p. 573-588.

FUSAR-POLI, P., G. DESTE, R. SMIESKOVA, S. BARLATI, A.R. YUNG, O. HOWES, R.D. STIEGLITZ, A. VITA, P. MCGUIRE et S. BORGWARDT (2012). « Cognitive functioning in prodromal psychosis : A meta-analysis », *Archives of General Psychiatry*, vol. 69, no 6, p. 562-571.

GEDDES, J., N. FREEMANTLE, P. HARRISON, P. BEBBINGTON (2000). « Atypical antipsychotics in the treatment of schizophrenia : Systematic overview and meta-regression analysis », *Archives of General Psychiatry*, vol. 321, no 7273, p. 1371-1376.

GREEN, M.F., K.H. NUECHTERLEIN, J.M. GOLD, D.M. BARCH, J. COHEN, S. ESSOCK, W.S. FENTON, F. FRESE, T.E. GOLDBERG, R.K. HEATON, R.S. KEEFE, R.S. KERN, H. KRAEMER, E. STOVER, D.R. WEINBERGER, S. ZALCMAN et S.R. MARDER (2004). « Approaching a consensus cognitive battery for clinical trials in schizophrenia : The NIMH-MATRICS conference to select cognitive domains and test criteria », *Biological Psychiatry*, vol. 56, nº 5, p. 301-307.

HAMILTON, M. (1960). « A rating scale for depression », *Journal of Neurology, Neurosurgery and Psychiatry*, vol. 23, p. 56-62.

HIGGINS, J.P., S.G. THOMPSON, J.J. DEEKS et D.G. ALTMAN (2003). « Measuring inconsistency in meta-analyses », *British Medical Journal*, vol. 327, nº 7414, p. 557-560.

HOWES, O.D., J. KAMBEITZ, E. KIM, D. STAHL, M. SLIFSTEIN, A. ABI-DARGHAM et S. KAPUR (2012). « The nature of dopamine dysfunction in schizophrenia and what this means for treatment », *Archives of General Psychiatry*, vol. 69, nº 8, p. 776-786.

ICHIKAWA, J. et H.Y. MELTZER (1999). « Relationship between dopaminergic and serotonergic neuronal activity in the frontal cortex and the action of typical and atypical antipsychotic drugs », *European Archives of Psychiatry and Clinical Neurosciences*, vol. 249, Suppl 4, p. 90-98.

LEANDRO, G. (2005). *Meta-Analysis in Medical Research : The Handbook for the Understanding and Practice of Meta-Analysis*, Londres, BMJ Books.

LEON, J. DE et F.J. DIAZ (2005). « A meta-analysis of worldwide studies demonstrates an association between schizophrenia and tobacco smoking behaviors », *Schizophrenia Research*, vol. 76, nºs 2-3, p. 135-157.

LEUCHT, S., C. CORVES, D. ARBTER, R.R. ENGEL, C. LI et J.M. DAVIS (2009). « Second-generation versus first-generation antipsychotic drugs for schizophrenia : A meta-analysis », *Lancet*, vol. 373, nº 9657, p. 31-41.

MACLEOD, C.M. (1991). « Half a century of research on the Stroop effect : An integrative review », *Psychological Bulletin*, vol. 109, nº 2, p. 163-203.

MARSMAN, A., M.P. VAN DEN HEUVEL, D.W. KLOMP, R.S. KAHN, P.R. LUIJTEN et H.E. HULSHOFF POL (2013). « Glutamate in schizophrenia : A focused review and meta-analysis of ^{1}H-MRS studies », *Schizophrenia Bulletin*, vol. 39, nº 1, p. 120-129.

MELTZER, H.Y. (1999). « The role of serotonin in antipsychotic drug action », *Neuropsychopharmacology*, vol. 21, nº 2, p. 106S-115S.

MOORE, T.H.M., S. ZAMMIT, A. LINGFORD-HUGHES *et al.* (2007). « Cannabis use and risk of psychotic or affective mental health outcomes : A systematic review », *Lancet*, vol. 370, p. 319-328.

MÖSSLER, K., X. CHEN, T.O. HELDAL et C. GOLD (2011). « Music therapy for people with schizophrenia and schizophrenia-like disorders », *Cochraine Database of Systematic Reviews*, vol. 7, nº 12, CD004025.

NIELSSEN, O.B., G.S. MALHI, P.D. MCGORRY et M.M. LARGE (2012). « Overview of violence to self and others during the first-episode of psychosis », *Journal of Clinical Psychiatry*, vol. 73, nº 5, p. 580-587.

POTVIN, S., A.A. SEPEHRY et E. STIP (2007). « A meta-analysis of depressive symptoms in dual-diagnosis schizophrenia », *Australian and New Journal of Psychiatry*, vol. 41, nº 10, p. 792-799.

POTVIN, S., K. STAVRO et J. PELLETIER (2012). « Paradoxical cognitive capacities in dual diagnosis schizophrenia : The quest for explanatory factors », *Journal of Dual Diagnosis*, vol. 8, nº 1, p. 35-47.

POTVIN, S., C.C. JOYAL, J. PELLETIER et E. STIP (2008a). « Contradictory cognitive capacities among substance-abusing patients with schizophrenia : A meta-analysis », *Schizophrenia Research*, vol. 100, nºs 1-3, p. 242-251.

POTVIN, S., K. STAVRO, E. RIZKALLAH et J. PELLETIER (soumis). « Cocaine and cognition : A systematic quantitative review », *Psychopharmacology*.

POTVIN, S., E. STIP, A.A. SEPEHRY, A. GENDRON, R. BAH et E. KOUASSI (2008b). « Inflammatory cytokine alterations in schizophrenia : A systematic quantitative review », *Biological psychiatry*, vol. 63, nº 8, p. 801-808.

RODER, V., D.R. MUELLER et S.J. SCHMIDT (2011). « Effectiveness of integrated psychological therapy (IPT) for schizophrenia patients : A research update », *Schizophrenia Bulletin*, vol. 37, nº 2, p. S71-S79.

RUMMEL, C., W. KISSLING et S. LEUCHT (2005). « Antidepressants as add-on treatment to antipsychotics for people with schizophrenia and pronounced negative symptoms : A systematic review of randomized trials », *Schizophrenia Research*, vol. 80, nº 1, p. 85-97.

RUMMEL-KLUGE, C., K. KOMOSSA, S. SCHWARZ, H. HUNGER, F. SCHMID, C.A. LOBOS, W. KISSLING, J.M. DAVIS et S. LEUCHT (2010). « Head-to-head comparisons of metabolic side effects of second generation antipsychotics in the treatment of schizophrenia : A systematic review and meta-analysis », *Schizophrenia Research*, vol. 123, nºs 2-3, p. 225-233.

SAVLA, G.N., L. VELLA, C.C. ARMSTRONG, D.L. PENN et E.W. TWAMLEY (2013). « Deficits in domains of social cognition in schizophrenia : A meta-analysis of the empirical evidence », *Schizophrenia Bulletin*, Epub non imprimé.

SEMPLE, D.M., A.M. MCINTOSH et S.M. LAWRIE (2005). « Cannabis as a risk factor for psychosis : Systematic review », *Journal of Psychopharmacology*, vol. 19, nº 2, p. 187-194.

SEPEHRY, A.A., S. POTVIN, R. ÉLIE et E. STIP (2007). « Selective serotonin reuptake inhibitor (SSRI) add-on therapy for the negative symptoms of schizophrenia : A meta-analysis », *Journal of Clinical Psychiatry*, vol. 68, nº 4, p. 604-610.

STAVRO, K., J. PELLETIER et S. POTVIN (2013). « Widespread and sustained cognitive deficits in alcoholism : A meta-analysis », *Addiction Biology*, vol. 18, nº 2, p. 203-213.

WYKES, T., V. HUDDY, C. CELLARD, S.R. MCGURK et P. CZOBOR (2011). « A meta-analysis of cognitive remediation for schizophrenia : Methodology and effect sizes », *American Journal of Psychiatry*, vol. 168, nº 5, p. 472-485.

CHAPITRE 9

L'ÉVALUATION D'IMPLANTATION DES PROGRAMMES
Illustration des programmes de réadaptation en santé mentale

Marie-José Durand
Marie-France Coutu
Quan Nha Hong

FORCES

- Elle permet de mieux comprendre les composantes actives d'un programme.
- Elle permet d'améliorer l'utilité des résultats en contribuant à définir des critères d'efficacité pertinents et en clarifiant les liens entre les processus et les résultats obtenus.
- Elle permet d'augmenter la validité externe en ciblant les milieux les plus susceptibles de bénéficier d'un programme.

LIMITES

- Elle requiert du temps et des ressources.
- Elle requiert des habiletés à bien comprendre les liens entre les objectifs et les activités pour construire un programme approprié.
- Encore peu d'études et de fonds de recherche sont disponibles pour ce type d'évaluation.

L'évaluation d'implantation (*implementation evaluation*), aussi connue sous l'appellation évaluation du processus (*process evaluation*), appartient au domaine de la recherche évaluative et vise à analyser les relations entre un programme et son contexte d'implantation dans la production des effets (Champagne *et al.*, 2011). Ce type d'évaluation permet de répondre à diverses questions portant sur le programme (p. ex., « Quelles sont les composantes du programme ? », « Quelles sont les forces et les limites du programme ? »), les personnes impliquées (p. ex., « Qui a participé au programme ? » « Qu'ont-elles fait ? ») et les facteurs contribuant à l'implantation d'un programme (Linnan et Steckler, 2002 ; Patton, 2008). L'évaluation d'implantation permet aussi d'étudier si un programme fonctionne comme initialement planifié. Ces types de questions sont complémentaires à celles étudiées dans l'évaluation des effets (p. ex., « Est-ce que le programme fonctionne ? ») et sont essentielles pour mieux comprendre avec quelles composantes, pour qui et sous quelles conditions un programme est efficace (Linnan et Steckler, 2002).

Ce chapitre est une introduction à l'évaluation d'implantation. D'abord, la raison d'être et l'utilité de l'évaluation d'implantation seront présentées ainsi que les approches d'évaluation d'implantation. Par la suite, des aspects pragmatiques de l'évaluation d'implantation seront présentés, c'est-à-dire les personnes impliquées, les moments opportuns et les méthodes de recherche à privilégier. Enfin, une dernière partie portera sur un exemple d'évaluation d'implantation dans le domaine de la réadaptation en santé mentale.

1. POURQUOI FAIRE UNE ÉVALUATION D'IMPLANTATION ?

Dans le domaine de la recherche évaluative en santé, et particulièrement en réadaptation, les principales préoccupations ont été jusqu'à maintenant de réaliser des analyses des effets pour savoir si un programme est efficace ou non. L'évaluation des effets s'appuie sur le modèle de la boîte noire où le programme est traité comme une variable dichotomique (c.-à-d. absence ou présence du programme) et est considérée comme une entité relativement homogène (Champagne *et al.*, 2011 ; Love, 2004 ; Patton, 2008). Dans un tel modèle, il est présumé que le programme présente peu ou pas de variations au moment de sa mise en œuvre et qu'il est imperméable à l'influence du contexte. Ainsi, les effets (ou l'absence d'effets) observés sont jugés sans être mis en relation avec les processus ayant contribué à les produire (Champagne *et al.*, 2011 ; Rossi, Lipsey et Freeman, 2004). Or les programmes en réadaptation sont souvent complexes. Un programme complexe regroupe un grand nombre d'activités, mobilisant de nombreux

acteurs interdépendants et ayant un certain degré de dépendance avec l'environnement (Contandriopoulos *et al.*, 2000). Il va de soi qu'une mise en œuvre inadéquate ou incomplète des éléments d'un programme peut limiter la portée des effets escomptés. Ceci fait référence à une erreur de type III, qui consiste à conclure qu'un programme est non efficace alors qu'il n'a pas été adéquatement implanté (Dobson et Cook, 1980). De plus, l'analyse des effets ne permet pas de bien comprendre ce qui fonctionne à l'intérieur d'un programme. Les effets documentés sont habituellement réservés aux résultats ultimes recherchés comme la diminution des symptômes, l'amélioration de la qualité de vie, le retour au travail ou encore la satisfaction des patients. Ainsi, les résultats à plus court terme ne sont pas souvent documentés et, par conséquent, ceci limite la compréhension de ce qui se passe pendant le programme ou à l'intérieur de celui-ci. L'évaluation des effets ne permettra donc pas de répondre à la question « Qu'est-ce qui fonctionne dans le programme ? » (Patton, 2008). Par conséquent, la réplication des programmes et leur évaluation sont limitées lorsqu'il y a seulement une évaluation des effets ultimes de ceux-ci (Patton, 2008 ; Porteous, 2009 ; Rossi *et al.*, 2004).

Pour contrer ces difficultés, de nombreux chercheurs ont plaidé pour la nécessité d'une évaluation d'implantation du programme préalablement à l'évaluation des effets (Champagne *et al.*, 2011 ; Dobson et Cook, 1980 ; Fixsen *et al.*, 2005 ; Linnan et Steckler, 2002 ; Patton, 2008 ; Rossi *et al.*, 2004). Love (2004) propose d'adopter le modèle de la boîte transparente, qui se base sur une perspective écologique dans laquelle les éléments du contexte organisationnel ainsi que la nature dynamique et interactive des programmes sont pris en compte. Dans cette perspective, on reconnaît qu'un programme avec de multiples composantes peut être influencé par son environnement. Dans le modèle de la boîte transparente, la façon dont un programme est mis en œuvre et les relations entre la théorie sous-jacente, les activités du programme et les résultats attendus sont explicités.

L'évaluation d'implantation revêt plusieurs utilités, qui peuvent être résumées en quatre grandes catégories. Premièrement, l'évaluation d'implantation permet d'améliorer la qualité d'un programme. En effet, une meilleure compréhension de la théorie sous-jacente permet de mieux comprendre les liens entre les processus d'un programme et son contexte (Bliss et Emshoff, 2002). Aussi, l'évaluation d'implantation permet de documenter les différentes composantes d'un programme, si ces composantes sont fonctionnelles et si le programme opère comme il est censé le faire (Patton, 2008). Ceci permettra de raffiner les composantes d'un programme et de connaître les forces et les limites de celui-ci dans un contexte réel

(Patton, 2008). Lorsqu'utilisée pour le monitorage, l'évaluation d'implantation assure l'amélioration continue du programme en fournissant rapidement des données sur les opérations effectuées et les résultats obtenus (Love, 2004).

Deuxièmement, l'évaluation d'implantation permet d'améliorer l'évaluation des effets d'un programme. En effet, elle permet de définir les variables indépendantes qu'il faut évaluer lors de l'évaluation des effets (Champagne *et al.*, 2011 ; Oakley *et al.*, 2006). Aussi, l'évaluation d'implantation permet de clarifier et de mieux interpréter les résultats obtenus dans l'évaluation des effets. Elle permet aussi de distinguer les programmes mal conceptualisés (cerner les problèmes de théorisation) de ceux qui sont mal implantés (cerner les problèmes d'implantation) (Oakley *et al.*, 2006).

Troisièmement, l'évaluation d'implantation permet de bien documenter comment les ressources ont été utilisées et ainsi de justifier leurs coûts. Ceci permet de rendre des comptes aux organismes ou individus qui subventionnent le programme (Love, 2004 ; Patton, 2008 ; Rossi *et al.*, 2004). De plus, elle peut fournir aux gestionnaires du programme des informations sur la qualité de l'implantation (Bliss et Emshoff, 2002). À l'issue d'une évaluation d'implantation, des décisions peuvent être prises sur la poursuite, l'expansion, la compression ou la cessation d'un programme.

Quatrièmement, les leçons tirées de l'implantation d'un programme peuvent servir à améliorer son implantation future dans d'autres sites (Patton, 2008). Généralement, trois principales raisons peuvent expliquer l'échec d'une implantation : 1) le programme n'a pas été correctement implanté ou seulement quelques composantes ont été implantées ; 2) un autre programme que celui visé a été implanté ; et 3) le programme n'est pas clairement défini et ni appliqué de façon similaire par les différents intervenants impliqués (Rossi *et al.*, 2004). Dans ces derniers cas, des correctifs peuvent être apportés afin d'assurer que le programme soit implanté avec succès dans d'autres sites.

1.1. Quelles sont les approches d'évaluation d'implantation ?

Plusieurs auteurs proposent diverses approches d'évaluation d'implantation (Champagne *et al.*, 2011 ; Chen, 2005 ; Patton, 2008 ; Provus, 1971). Ces approches se distinguent par leurs appellations, leurs objectifs, les populations visées ou le stade d'implantation. Au Québec, une typologie a été développée par des chercheurs de l'Institut de recherche en santé publique de l'Université de Montréal et est enseignée dans les institutions universitaires. Trois types d'évaluation d'implantation y sont définis (Champagne *et al.*, 2011), présentés dans le tableau 9.1.

Tableau 9.1.
Types d'évaluation d'implantation

Type d'évaluation d'implantation	Visée	Exemples de questions
Type 1a	Documenter comment le contexte influence le programme et comment le programme se transforme au cours du temps.	Quelles sont les composantes du programme ? Quels facteurs du contexte influencent la mise en œuvre du programme ?
Type 1b	Analyser le degré de mise en œuvre d'un programme.	Quel est l'écart entre ce qui était initialement prévu et ce qui est réellement implanté ?
Type 2	Analyser l'influence de la variation d'implantation sur les effets observés.	Quelles sont les variations dans les résultats observés ? Quels sont les aspects de la mise en œuvre qui facilitent ou entravent l'atteinte des résultats ?
Type 3	Étudier l'interaction entre le contexte et l'intervention sur les effets observés.	Quels facteurs contextuels facilitent ou entravent les effets observés ?

Source : Adapté de Champagne *et al.*, 2011.

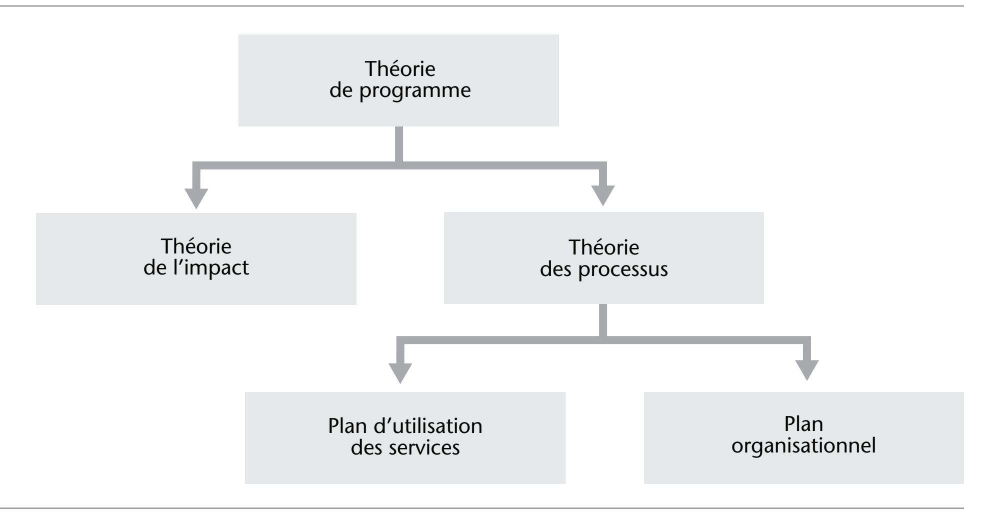
Par conséquent, avant d'amorcer une évaluation, il est essentiel de bien cibler le type d'évaluation d'implantation qui sera réalisée, car les méthodes et l'ampleur du travail vont varier considérablement de l'une à l'autre.

Une des premières étapes en évaluation de programme est de rendre explicite la théorie de programme. Le développement d'une théorie de programme est important pour mieux comprendre un programme, faciliter son implantation, structurer l'évaluation et orienter les questions de recherche. La *théorie de programme* (*program theory*) décrit comment un programme devrait fonctionner pour obtenir les résultats visés (Rossi *et al.*, 2004). Comme présenté à la figure 9.1, la théorie de programme comporte deux composantes, soit la théorie des processus (*process theory*) et la théorie de l'impact (*impact theory*) (Rossi *et al.*, 2004).

La *théorie de l'impact* concerne la séquence des hypothèses causales qui vont entraîner les changements désirés (Rossi *et al.*, 2004). La *théorie des processus* reflète l'ensemble des ressources (structure) et des activités (processus) impliquées pour atteindre les changements désirés. Cette théorie est composée du plan d'utilisation des services et du plan organisationnel. Le *plan d'utilisation des services* (*service utilization plan*) décrit les différentes étapes du programme qu'un participant aura à suivre, du premier contact à

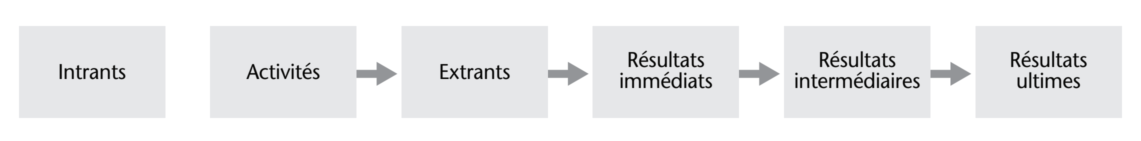
la fin du programme. Le *plan organisationnel* (*organizational plan*) est surtout utilisé par les gestionnaires du programme et décrit les ressources (humaines, financières et physiques) nécessaires pour offrir les activités prévues afin d'atteindre les résultats visés (Rossi *et al.*, 2004).

Figure 9.1.
Théorie de programme



La théorie de programme peut être représentée graphiquement par ce qui est communément appelé un *modèle logique* (McLaughlin et Gretchen, 2010; Porteous, 2009). Dans un modèle logique, les ressources (intrants), les activités (processus), les produits ou services découlant des activités (extrants) et les résultats immédiats (à court terme), intermédiaires (à moyen terme) et ultimes (à long terme) d'un programme sont généralement présentés dans des cases reliées par des flèches (figure 9.2). Différents exemples de modèles logiques en réadaptation sont disponibles (Équipe CREXE, 2010; Letts *et al.*, 1999; Pauls, 2004; Sly et Taylor, 2005).

Figure 9.2.
Éléments de base du modèle logique



1.2. Quelles sont les personnes impliquées dans l'évaluation d'implantation ?

Trois principaux groupes de personnes sont impliqués en évaluation de programme. Un premier groupe de personnes sont les évaluateurs qui peuvent être internes (c.-à-d. faisant partie de l'organisation où a lieu l'évaluation) ou bien externes. L'évaluateur interne peut avoir plus de facilité à développer une connaissance détaillée du programme alors que celui qui est externe au programme peut faire preuve de plus d'objectivité (Nadeau, 1988). Un évaluateur peut également être amené à jouer d'autres rôles comme ceux de concepteur du programme, de formateur, de conciliateur, de facilitateur, d'agent de changement, de négociateur et de consultant (Chen, 2005; Nadeau, 1988). Lorsqu'un évaluateur possède plusieurs rôles, il est important de porter attention au conflit d'intérêts possible pouvant influencer la crédibilité des résultats. Par exemple, il pourrait être jugé inadéquat qu'un évaluateur ayant participé de façon active au développement d'un programme dans ses phases préliminaires soit également responsable d'évaluer son implantation et son efficacité. Dans ce cas, l'évaluateur pourrait avoir un parti pris pour le programme et ne pas être en mesure de porter un regard objectif lors de l'évaluation (Chen, 2005).

Un second groupe est composé du personnel qui offre le programme (les fournisseurs de services) ainsi que du personnel du soutien administratif et de la haute direction. Ces personnes peuvent jouer un rôle important dans l'évaluation, car elles sont directement impliquées dans l'implantation du programme. Elles peuvent contribuer à l'élaboration du modèle logique et de la théorie de programme, au recueil des données pour le monitorage du programme et rapporter des difficultés d'implantation du programme. Leur participation à l'évaluation est donc essentielle pour fournir des informations sur les ressources humaines, financières et matérielles requises et pour mieux comprendre comment le programme fonctionne.

Enfin, les participants (utilisateurs de services) sont les personnes qui reçoivent le programme. Leur participation à l'évaluation permet de s'assurer que le programme a bien été offert comme planifié. Ceci peut être fait en documentant les activités et les services offerts à ces participants ainsi que les délais pour offrir les services. Aussi, documenter les caractéristiques des participants permet de s'assurer que le programme rejoint bien la population ciblée. En réadaptation, d'autres personnes que les utilisateurs de services peuvent aussi être sollicitées dans le cadre d'un programme. Il peut s'agir, par exemple, des parents d'un enfant, de l'employeur d'un travailleur ou du proche aidant d'un aîné. La participation de ces personnes au processus d'évaluation pourrait être sollicitée en particulier lorsqu'elles jouent un rôle actif dans le processus et influencent le succès ou non d'un programme.

1.3. Quand faire une évaluation d'implantation ?

Les chercheurs s'entendent pour dire que l'évaluation d'implantation devrait être faite avant l'évaluation des effets (Chen, 2005 ; Oakley *et al.*, 2006 ; Patton, 2008). Il n'est toutefois pas rare de voir ce type d'évaluation menée en parallèle à l'évaluation des effets afin de définir comment les programmes sont reliés aux résultats visés. L'évaluation d'implantation devrait être privilégiée lorsque le programme est complexe, n'est pas standard et est sensible au contexte dans lequel il est implanté (Oakley *et al.*, 2006). De plus, ce type d'évaluation est de mise lorsqu'un programme est implanté dans différents sites puisque l'implantation d'une même intervention pourrait prendre différentes formes (Oakley *et al.*, 2006). Aussi, Champagne *et al.* (2011) précisent que l'évaluation d'implantation est utile lorsque des effets différents sont observés selon le contexte.

Chen (2005) distingue des types d'évaluation selon le stade d'implantation d'un programme. Pour une implantation récente, il propose d'utiliser une approche d'évaluation plus souple afin d'obtenir des résultats plus rapidement et d'ajuster le programme et les stratégies d'implantation en conséquence. L'évaluation portera surtout sur la manière dont le programme a été conceptualisé et sur les problèmes d'implantation. Pour les programmes dont l'implantation est de plus longue durée, il propose d'utiliser une approche plus rigoureuse afin d'évaluer la qualité de l'implantation et d'établir si le programme implanté diffère de celui initialement planifié. L'évaluation à ce stade, qu'il nomme évaluation de processus (*process evaluation*), requiert généralement plus de temps afin de fournir des données rigoureuses sur le succès de l'implantation d'un programme.

Une autre distinction sur le moment de l'évaluation concerne l'évaluation continue par opposition à l'évaluation ponctuelle. Pour des programmes qui sont implantés depuis longtemps, le monitorage de ceux-ci est nécessaire pour s'assurer que les services offerts restent cohérents avec leurs objectifs et qu'ils répondent aux besoins de la population ciblée (Rossi *et al.*, 2004). Pour ce faire, des indicateurs sur les opérations (p. ex., nombre de participants évalués, nombre d'offres de service), les ressources (p. ex., nombre d'intervenants impliqués, coûts du programme) et la performance du programme (p. ex., délai d'attente avant de recevoir un traitement, durée du programme) sont colligés sur une base continue.

1.4. Quels sont les devis et méthodes à privilégier en évaluation d'implantation ?

Le devis de recherche sera différent selon le type d'évaluation d'implantation (1, 2 ou 3). Dans le cas d'une évaluation de type 1 (voir le tableau 9.1), l'étude descriptive simple[1] est généralement utilisée. Habituellement, à partir de la théorie de programme et des objectifs visés, des indicateurs sont créés, c'est-à-dire des informations à colliger pour objectiver une réalité (Jabot et Bauchet, 2009). Ces indicateurs permettent de suivre l'ensemble du programme ainsi que de valider les caractéristiques de la population desservie. Des exemples d'indicateurs sont présentés dans le tableau 9.2 dans la section suivante sur l'illustration d'une évaluation d'implantation. Ainsi, l'écart entre ce qui était prévu et la réalité peut facilement être établi.

Pour les évaluations de type 2 ou 3 (voir le tableau 9.1), l'étude de cas simple ou multiples est privilégiée (voir le chapitre 4 du présent ouvrage). Les études de cas simple ou multiples permettent de comprendre plus en profondeur l'implantation du programme en ajoutant une analyse des facteurs qui expliquent les variations possibles lors de celle-ci. Dans ces types d'évaluation d'implantation, il ne s'agira pas seulement de générer des indicateurs, mais d'utiliser d'autres méthodes telles que l'entrevue pour aller chercher le sens inhérent à l'expérience vécue des participants (utilisateurs de service) et des fournisseurs de services.

Généralement, plusieurs méthodes de collecte de données sont utilisées. Il peut s'agir d'un tableau de bord avec des indicateurs qui comprend des variables nominales (p. ex., présence ou absence d'une évaluation psychiatrique) ou quantitatives (p. ex., nombre d'heures de thérapie par semaine). D'autres méthodes comme des entrevues ou des groupes de discussion avec des guides d'entretien, des observations non participantes et des sondages peuvent également être utilisées. Le choix des méthodes de recherche dépendra de plusieurs critères, dont la nature du programme, l'objectif de l'étude, les personnes concernées, la disponibilité des données, les documents existants, les systèmes en place, le nombre et la dispersion des sites d'implantation. Il est aussi important de considérer le temps, les ressources et le financement disponibles accordés aux chercheurs.

1. Le devis descriptif simple consiste à « décrire les caractéristiques, les comportements et les conditions de personnes ou de groupes » (Fortin, 2010, p. 292).

2. ILLUSTRATION D'UNE ÉVALUATION DE L'IMPLANTATION D'UN PROGRAMME DE RETOUR AU TRAVAIL POUR DES PERSONNES PRÉSENTANT DES TROUBLES MENTAUX TRANSITOIRES

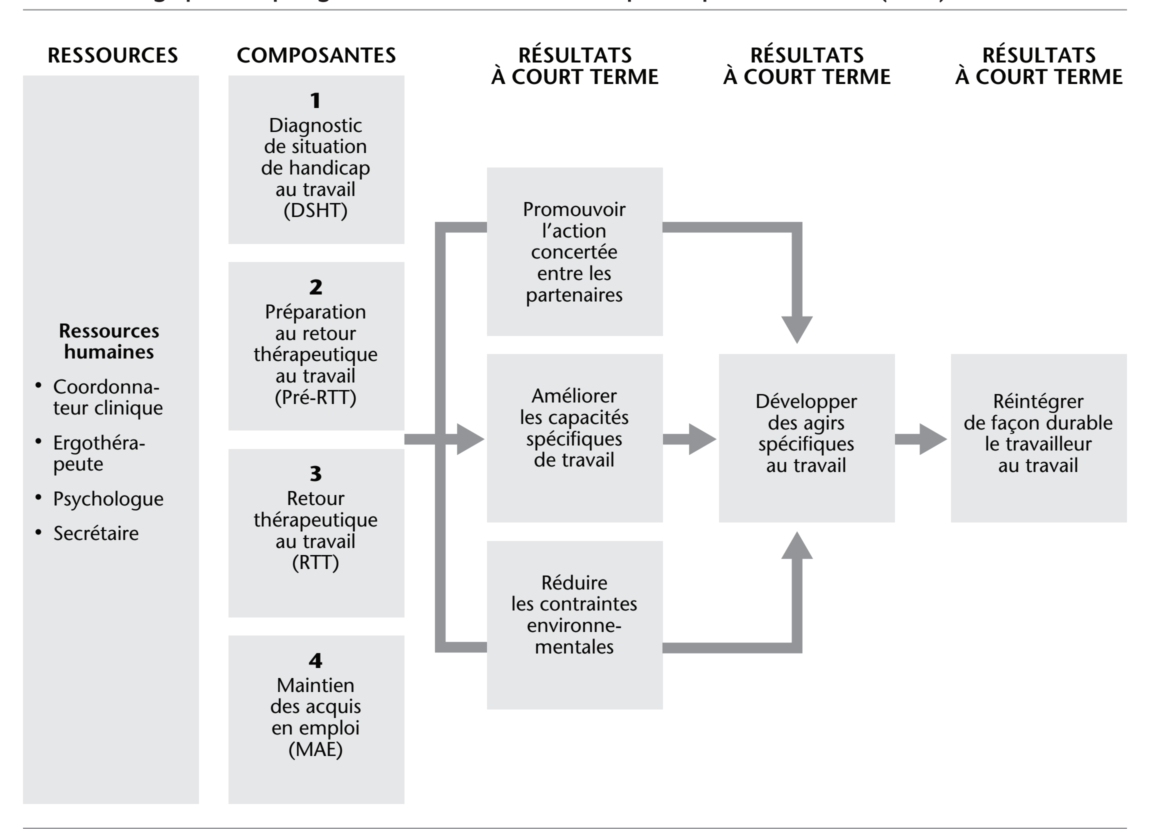
Dans cette section, un exemple d'étude d'implantation d'un programme visant le retour au travail de travailleurs ayant un trouble mental transitoire (voir le chapitre 6 de cet ouvrage pour la définition du trouble mental transitoire) sera présenté. Il s'agit d'une étude pilote effectuée auprès de travailleurs d'une même organisation. Le programme en question, le programme de retour thérapeutique au travail (RTT), a été initialement développé à la fin des années 1990 pour des travailleurs ayant des troubles musculosquelettiques. Dans leur étude qualitative, Briand, Durand, St-Arnaud et Corbière (2007) ont clairement établi des ponts entre ce programme de RTT conçu pour des personnes présentant des troubles musculosquelettiques avec celui qui pourrait être offert à des personnes souffrant de troubles mentaux transitoires en absence du travail. Cependant, aucune application officielle n'avait été réalisée jusqu'à ce jour. L'objectif de ce projet pilote était donc de tester la faisabilité du programme de RTT auprès de travailleurs ayant un trouble mental transitoire.

2.1. Programme

Le modèle logique du programme de RTT est présenté à la figure 9.3. Les ressources humaines requises pour ce programme sont un coordonnateur clinique, un ergothérapeute, un psychologue et une secrétaire. Les membres de ce programme partagent une vision commune en réadaptation au travail, prônent l'approche centrée sur le client et sont orientés vers la recherche de solutions. Quatre principales composantes sont définies dans le programme de RTT. Il s'agit 1) du diagnostic de situation de handicap au travail (DSHT), dont le but principal est d'éliminer un diagnostic de gravité et de cerner la ou les causes de la prolongation d'une incapacité au travail ; 2) de la préparation au retour thérapeutique au travail (pré-RTT), qui vise la mise en place en milieu clinique des conditions favorables qui permettent la réintégration en milieu de travail ; 3) du retour thérapeutique au travail (RTT), qui consiste en une reprise progressive du travail supervisée au poste régulier ou à un poste semblable, d'abord avec des tâches allégées, puis une augmentation progressive des contraintes, des cadences et de la durée du travail en fonction des capacités du travailleur ; et 4) du maintien des acquis en emploi (MAE), qui vise, par souci de prévention d'une rechute, la stabilité et le maintien des capacités développées lors du RTT ainsi que la mise en place d'une réduction des contraintes environnementales (Durand et Loisel, 2001a, 2001b ; Durand, Loisel, Charpentier, Labelle et Hong, 2004 ; Durand, Loisel, Hong et Charpentier, 2002 ; Durand, Vachon, Loisel et

Berthelette, 2003). Des réunions hebdomadaires de l'équipe clinique sont réalisées afin d'établir les objectifs de traitement. Aussi, pendant les composantes 2 et 3 (pré-RTT et RTT), des entretiens et des contacts téléphoniques réguliers auprès du conseiller de l'assureur, de l'employeur et du médecin traitant du travailleur sont effectués au besoin. Les objectifs visés par le programme sont de promouvoir l'action concertée entre les partenaires, d'améliorer les capacités spécifiques de travail et de réduire les contraintes environnementales afin de développer des agirs spécifiques au travail, c'est-à-dire d'adopter en milieu réel de travail les comportements attendus pour la réalisation des tâches de travail. Ces objectifs visent ultimement à favoriser la réintégration de façon durable au travail. Notez que dans le modèle logique présenté à la figure 9.3, les extrants n'ont pas été ajoutés puisqu'ils sont nombreux. Les indicateurs présentés dans le tableau 9.2 peuvent être considérés comme des extrants dans le cadre de cette étude.

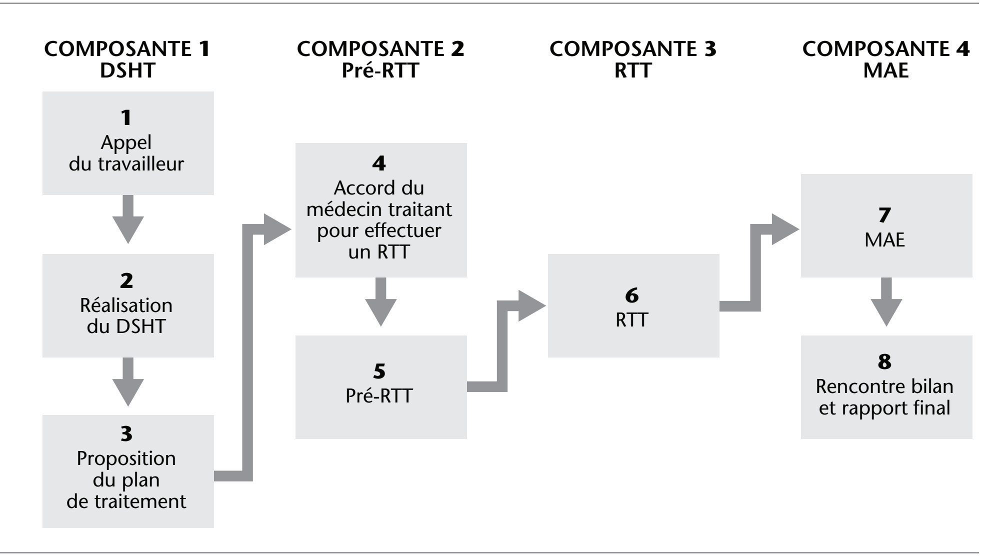
Figure 9.3.
Modèle logique du programme de retour thérapeutique au travail (RTT)



La théorie du programme a été développée à la fois par le concepteur du programme de RTT, les cliniciens offrant le programme et un chercheur en évaluation de programme. Elle a été produite en se basant sur l'analyse des documents décrivant l'approche privilégiée lors du RTT et particulièrement sur l'expérience et l'observation sur le terrain (Durand *et al.*, 2003). L'approche de RTT repose sur les principaux fondements de la théorie d'adaptation au travail développée par une équipe de chercheurs de l'Université du Minnesota (Dawis et Lofquist, 1984; Dawis, Lofquist et Weiss, 1968). Cette théorie considère l'adaptation au travail comme la résultante de l'interaction entre l'individu et les caractéristiques de son emploi. La théorie de programme du RTT a été évaluée au début des années 2000 dans le cadre d'une étude auprès de travailleurs dans le secteur de la construction (Durand *et al.*, 2012a, 2012b).

Dans le cadre de la présente étude, le plan d'utilisation des services original a été adapté afin de tenir compte du contexte et de la population de l'étude. Comme illustré à la figure 9.4, le plan comprend huit étapes séquentielles.

Figure 9.4.
Plan d'utilisation des services du programme de RTT



2.2. Objectifs

Dans l'objectif général de tester la faisabilité du programme de RTT pour les personnes ayant un trouble mental transitoire, une étude pilote a été menée auprès de travailleurs d'une organisation. Les objectifs spécifiques de l'étude étaient 1) d'apprécier le degré de mise en œuvre du programme de RTT en comparant les ressources et les composantes du programme offertes avec celles du plan d'utilisation théorique (évaluation de type 1b) ; et 2) de déterminer quels facteurs contextuels expliquent la variation de l'implantation du programme (évaluation de type 1a).

2.3. Méthode

2.3.1. Devis de recherche

Le devis retenu était une étude de cas multiples (Yin, 2003), car ce devis permet l'étude en profondeur d'un petit nombre de cas (voir le chapitre 4 de cet ouvrage). Chaque cas était défini par le travailleur absent du travail, son supérieur immédiat, les cliniciens et le coordonnateur du service d'assurance. Les critères d'inclusion des travailleurs étaient les suivants : 1) être en absence du travail pour un trouble mental transitoire (dans cette étude, les diagnostics suivants étaient considérés : trouble de l'adaptation, dépression majeure, trouble d'anxiété généralisée ou trouble panique avec agoraphobie) ; 2) être en absence du travail depuis au moins trois mois ; 3) être âgé entre 18 et 64 ans ; et 4) accepter de divulguer son trouble mental transitoire au supérieur immédiat.

2.3.2. Recrutement des participants

L'organisation participante dans cette étude était une agence gouvernementale provinciale. Cette organisation comptait plus de 6000 employés et observait un taux important d'absentéisme, dont plus de la moitié était liée à des troubles mentaux transitoires. Le coordonnateur du service d'assurance de cette organisation était responsable de l'envoi des lettres aux travailleurs admissibles à l'étude. Cet envoi était constitué de deux lettres. La première informait les travailleurs de la collaboration de l'organisation avec une équipe de recherche. Une deuxième lettre du centre de recherche informait les travailleurs qu'ils étaient admissibles et libres de participer à l'étude. Elle informait aussi les travailleurs du but de l'étude et de leur rôle en tant que participant. Cette méthode de recrutement a été retenue pour respecter la confidentialité des travailleurs et obtenir un consentement

volontaire de leur part. Les travailleurs intéressés à participer à l'étude contactaient l'équipe clinique et se voyaient offrir le programme de RTT. Avant de commencer la prise en charge, l'accord du médecin traitant devait être obtenu.

2.3.3. Collecte de données

Pour l'objectif spécifique 1, s'agissant de décrire la variation des ressources et des composantes utilisées dans le programme de RTT, des informations détaillées ont été recueillies lors de chaque étape du programme. Le tableau 9.2 présente les indicateurs choisis pour chacune des étapes. Ces informations sur les délais, les ressources humaines et les composantes ont été recueillies à partir des dossiers médicaux des sujets et auprès des cliniciens impliqués dans la prestation du programme, et ce, à partir d'un formulaire développé à cet effet.

Pour l'objectif spécifique 2, des entrevues individuelles semi-dirigées ont été effectuées à la fin du programme de RTT avec les travailleurs (utilisateurs de service) et les acteurs impliqués dans les démarches, c'est-à-dire les supérieurs immédiats, le coordonnateur du service d'assurance et les cliniciens (fournisseurs de service). Des guides d'entrevues ont été préalablement développés pour documenter les principales questions suivantes : « Est-ce que toutes les composantes du programme ont été implantées ? », « Sinon, qu'est-ce qui explique l'absence d'implantation de certaines composantes du programme ? », « Quels ont été les éléments facilitateurs ou les obstacles pour l'implantation du programme (p. ex., disponibilité des ressources, ouverture du milieu de travail, relation avec le supérieur immédiat et collègues, etc.) ? » Tous les participants ont signé un formulaire de consentement pour répondre aux normes du comité d'éthique et de la recherche de l'Hôpital Charles-Le Moyne.

Tableau 9.2.
Exemples d'indicateurs d'implantation correspondants aux étapes du programme présentées dans le plan d'utilisation des services (figure 9.4)

Étapes	Indicateurs
1. Appel du travailleur	▪ Date de l'appel du travailleur
2. Réalisation du DSHT	▪ Date du DSHT ▪ Disciplines impliquées (ergothérapie, ergonomie, psychologie) ▪ Consultations supplémentaires demandées (discipline, heures allouées)

(*suite*)

Tableau 9.2. (*suite*)

3. Proposition du plan de traitement	▪ Date de la proposition du plan de traitement ▪ Date de la réponse des partenaires ▪ Nature de l'orientation ▪ Réception de l'accord des partenaires (accepté, accepté avec ajustement, refusé par conseiller et médecin traitant, date)
4. Accord du médecin traitant	▪ Date d'envoi du rapport au médecin ▪ Date de la réponse du médecin ▪ Accord ou non ; si refus, motifs invoqués
5. Pré-RTT	▪ Date du début du pré-RTT ▪ Nombre de semaines en pré-RTT ▪ Disciplines impliquées (ergothérapie, ergonomie, psychologie) ▪ Heures allouées dans chaque discipline ▪ Contacts réalisés avec l'employeur (durée, fréquence, par qui, nature des informations recueillies, appel ou visite) ▪ Contacts réalisés avec l'assureur (durée, fréquence, par qui, nature des informations recueillies, appel ou visite) ▪ Contacts réalisés avec le médecin traitant (durée, fréquence, par qui, nature des informations recueillies, appel ou visite) ▪ Nombre de réunions interdisciplinaires
6. RTT	▪ Date du début du RTT ▪ Accord de l'employeur pour effectuer le RTT (accord ou non ; si refus, motifs invoqués) ▪ Nombre de semaines en RTT ▪ Disciplines impliquées (ergothérapie, ergonomie, psychologie) ▪ Heures allouées dans chaque discipline ▪ Contacts réalisés avec l'employeur (durée, fréquence, par qui, nature des informations recueillies, appel ou visite) ▪ Contacts réalisés avec l'assureur (durée, fréquence, par qui, nature des informations recueillies, appel ou visite) ▪ Contact réalisé avec le médecin traitant (durée, fréquence, par qui, nature des informations recueillies, appel ou visite) ▪ Nombre de réunions interdisciplinaires
7. MAE	Date du début du MAE Nombre de suivis Disciplines impliquées (ergothérapie, ergonomie, psychologie) Type de suivi (téléphonique, courriel, en personne) Nombre de réunions interdisciplinaires
8. Fin de programme	Date de fin de programme

2.3.4. Analyse des données

Toutes les entrevues (n = 27) ont été enregistrées sur bande audio et ont été transcrites sous forme de comptes rendus intégraux (*verbatim*) et anonymisées aux fins d'analyse. Le logiciel d'aide à l'analyse de contenu Atlas/ti 5.0 a été utilisé pour soutenir l'analyse des comptes rendus (*verbatim)* (Muhr, 2004). L'analyse du contenu des entrevues s'est fondée sur la méthode de Landry (1997). Il s'agit de cinq étapes, soit 1) la détermination des objectifs de l'analyse de contenu; 2) la préanalyse; 3) l'analyse du matériel étudié; 4) l'évaluation de la fiabilité et de la validité des données; et 5) l'analyse et l'interprétation des résultats. Ainsi, l'analyse de contenu des entrevues, qui se voulait thématique, s'est d'abord effectuée à partir d'une liste de thèmes provenant des écrits scientifiques sur les facilitants et les obstacles au retour au travail et également sur les écarts d'implantation de programmes complexes. Par exemple, certains thèmes retenus concernaient l'adhésion du participant, le rôle de l'assureur et l'action concertée entre les acteurs. Un premier guide d'analyse qui contenait ces thèmes a été élaboré par l'équipe de recherche puis prétesté de façon à s'assurer que les thèmes étaient clairs, pertinents, exclusifs et fidèles. Ces étapes ont été poursuivies jusqu'à l'obtention d'un niveau d'accord acceptable, soit de plus de 80 % entre trois codeurs. Les résultats divergents ont été discutés et clarifiés jusqu'à un niveau d'accord satisfaisant entre les codeurs. À la suite de la détermination du guide d'analyse avec une fidélité interjuges acceptable, les comptes rendus (*verbatim)* ont été analysés par un codeur. Cependant, une vérification a été réalisée à la moitié du codage de l'ensemble des entrevues. En effet, le second codeur a refait un nouveau codage d'une entrevue et a comparé les résultats au premier codage afin de s'assurer que le codage n'avait pas été modifié.

2.4. Résultats

Un total de huit travailleurs absents du travail en raison d'un trouble mental transitoire ont participé à cette étude, et 27 entrevues ont été effectuées auprès des travailleurs, cliniciens, supérieurs immédiats et coordonnateur du service d'assurance.

À titre de rappel, l'*objectif 1* était d'apprécier le degré de mise en œuvre du programme de RTT. Les données colligées sur les indicateurs préalablement retenus (voir le tableau 9.2) ont permis de comparer les ressources et les composantes du programme offertes avec celles théoriques (plan d'utilisation théorique). À titre d'exemple, nous présentons dans ce chapitre les délais de chacune des étapes du plan d'utilisation des services. Le tableau 9.3 compare les délais prévus avec les délais réels.

Tableau 9.3.
Délais moyens entre les étapes du plan d'utilisation de service théorique comparativement à ceux du plan offert

Étapes du plan d'utilisation des services	Théorique (jours)	Réel (jours) Moyenne (étendue)
Étape 1 : Appel du travailleur (délai entre l'absence du travail et l'appel du travailleur) (*n* = 8)	S. O.	147 (63–285)
Étape 2 : Réalisation du DSHT (délai entre l'appel du travailleur et la réalisation du DSHT) (*n* = 8)	7	12 (7–16)
Étape 3 : Proposition du plan de traitement (délai entre la réalisation du DSHT et la proposition du plan de traitement) (*n* = 8)	10	45 (16–115)
Étape 4 : Accord du médecin traitant (délai entre la proposition du plan de traitement et l'accord du médecin (*n* = 7)	10	–17 (–109 – 18)*
Étape 5 : Durée du pré-RTT (*n* = 6)	84 (étapes 5 à 7)	54 (0–116)
Étape 6 : Durée du RTT (*n* = 6)		106 (38–267)
Étape 7 : Durée du MAE (*n* = 4)		61 (0–196)

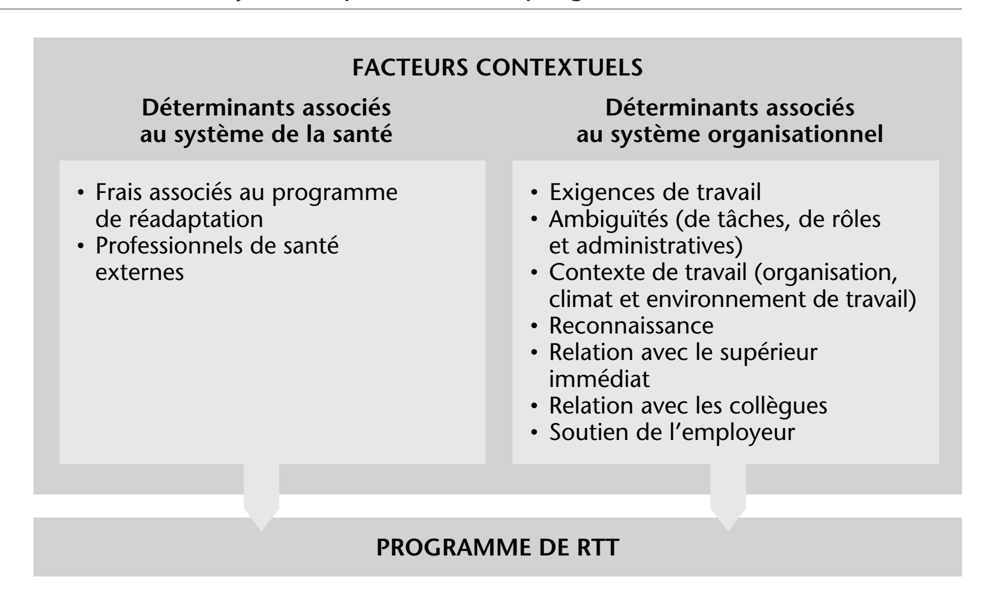
* La valeur négative indique des inversions dans la séquence des composantes.

À partir des données colligées sur les délais présentées dans le tableau 9.3, trois principaux constats peuvent être tirés. Premièrement, les délais réels étaient plus longs que ceux prévus (théorique). Par exemple, pour les étapes de pré-RTT, du RTT et du MAE (étapes 5 à 7), il avait été estimé que 12 semaines seraient nécessaires. Or la durée de la prise en charge s'est échelonnée sur 32 semaines en moyenne. Deuxièmement, il y a eu une variation sur l'offre des composantes du programme, à savoir : trois travailleurs ont reçu toutes les trois composantes après l'évaluation (étapes 5, 6 et 7), deux ont reçu le pré-RTT et le RTT (étapes 5 et 6), une personne a reçu le RTT et le MAE (étapes 6 et 7) et enfin une personne a reçu seulement le pré-RTT (étape 5) avant d'être retirée du programme en raison d'un problème de santé physique. Troisièmement, la séquence des étapes n'a pas été suivie. En effet, dans certains cas, l'accord du médecin traitant (étape 4) a été obtenu avant même la proposition du plan de traitement (étape 3). Ceci explique qu'il y a une valeur négative dans le tableau 9.3.

L'*objectif 2* visait à déterminer quels facteurs contextuels expliquent la variation de l'implantation du programme de RTT. Les entrevues auprès des travailleurs, cliniciens, supérieurs immédiats et coordonnateur du service d'assurance ont permis de mieux comprendre divers facteurs pouvant influencer l'implantation du programme. Comme il a été mentionné, toutes

les entrevues ont été codées avec la liste de thèmes et de codes prétestés. Ensuite, une analyse verticale (intracas) a été effectuée pour chaque cas et a permis de retenir les thèmes convergents et divergents entre les acteurs dans un cas. Enfin, une analyse transversale (intercas) a permis de révéler les thèmes communs entre les cas. La figure 9.5 présente les facteurs identifiés dans cette étude. En définitive, il est possible d'observer que ces facteurs sont en lien avec les composantes du programme, les caractéristiques des utilisateurs de service et différents systèmes (système de santé, système organisationnel). Aussi, certains facteurs pouvaient agir à la fois comme un obstacle et un facilitateur.

Figure 9.5.
Facteurs contextuels influençant l'implantation du programme



Cette étude a permis d'observer que la majorité des étapes du plan d'utilisation des services a été mise en œuvre pour la plupart des cas, mais que les moments de l'offre de certaines composantes ont été déplacés et que les délais étaient beaucoup plus longs que ceux prévus (grands écarts entre les délais théoriques et réels). Cet écart important en matière de délai et de durée de programme peut avoir des conséquences sur l'implantation du programme. En effet, comme ce programme requiert des ressources sur une plus longue période de temps, il peut être plus coûteux, ce qui limite ainsi sa réplication et son acceptabilité dans la pratique courante. Cette

étude a également permis de cerner comment les différents facteurs du milieu de travail (p. ex., le soutien de l'employeur, l'organisation des postes de travail) pouvaient faire basculer la démarche positivement ou négativement et entraîner une prolongation importante du programme de RTT. À la lumière de ce type de résultats, c'est-à-dire l'influence des facteurs organisationnels sur le niveau d'implantation d'un programme, les chercheurs ont pu établir assez clairement dans quelles conditions un tel programme devrait être offert. Par exemple, dans notre illustration, une condition essentielle pour faire de l'exposition au travail (RTT) était de s'assurer d'obtenir la collaboration avec l'employeur dès le début de la démarche. Encore une fois, ces résultats soulignent l'importance d'effectuer l'évaluation de l'implantation d'un programme avant l'évaluation de ses effets en tant que tels.

Malgré la richesse des données recueillies, l'évaluation d'implantation exige beaucoup de ressources et de temps afin de documenter l'ensemble des variables à l'étude pendant le déroulement du programme et à la fin de celui-ci. Plusieurs difficultés ont été surmontées lors de cette étude. Premièrement, la collecte de données avait lieu auprès de divers acteurs à des moments précis. Il était donc important de bien coordonner les différentes étapes du programme et les moments des entrevues et de collecte de données. Aussi, un faible taux de refus des autres acteurs impliqués (supérieur immédiat, représentant de l'assurance) est probablement dû au fait que l'étude s'est déroulée chez l'employeur, qui s'était activement impliqué dans l'étude. Le protocole a été présenté à tous les directeurs et aux représentants syndicaux.

Deuxièmement, plusieurs données temporelles auprès des cliniciens devaient être colligées. Afin d'éviter un biais de mémoire, un formulaire de collecte de données pour chacune des composantes du programme a été développé. Aussi, des rappels fréquents auprès des cliniciens étaient requis de la part du professionnel de recherche, qui assurait un suivi quotidien de la progression de chacun des cas. De plus, le bureau de ce professionnel de recherche était localisé à proximité de ceux de l'équipe clinique, ce qui facilitait le suivi et les rappels. En fait, plus la personne qui réalise la collecte de données est loin du contexte dans lequel s'offre le programme, plus des stratégies doivent être mises en œuvre pour éviter de perdre des données. Également, il faut toujours s'assurer, lorsque le programme est offert dans un contexte naturel, que les méthodes utilisées pour la collecte de données (p. ex., tableau de gestion d'indicateurs, journal de bord, fichier de saisie de données, etc.) sont simples, rapides et facilement accessibles. Ces dernières constituent souvent une surcharge de travail pour les personnes en place (fournisseurs de service). Il se révèle souvent éclairant de questionner les personnes en place pour trouver les meilleures

stratégies pour recueillir les informations. Par exemple, est-ce une feuille dans le dossier médical, une liste de rappels que suit la secrétaire de la clinique ou encore un courriel envoyé comme rappel qui sera la meilleure méthode ? Troisièmement, telle qu'observée dans cette étude, la durée de la prise en charge des patients était en moyenne de 32 semaines et cette étude s'est échelonnée sur 18 mois. Ce projet nécessitait donc un investissement temporel important pour un petit nombre de cas. Par conséquent, dans le cas où il s'agit d'une première implantation de programme, il faut estimer la durée de déploiement de celui-ci et prévoir des ressources à temps partiel sur une période prolongée. De plus, nous recommandons de faire une étude pilote auprès des premiers cas (deux à trois cas) avant d'amorcer l'étude d'implantation. Ceci permettra de rectifier les problèmes liés au démarrage et de procéder à l'ajustement initial du programme.

CONCLUSION

Ce chapitre visait à présenter ce qu'est l'évaluation d'implantation et à illustrer ce type d'étude à l'aide d'un exemple concret. En évaluation de programme, ce type d'évaluation est essentiel pour mieux comprendre comment un programme fonctionne, pour l'améliorer et pour favoriser son implantation dans d'autres sites. À l'heure actuelle, il n'existe pas de consensus sur la meilleure façon de procéder pour effectuer une évaluation d'implantation, chacune devant être adaptée au contexte de l'étude. L'évaluation d'implantation devrait être privilégiée pour les programmes complexes, ce qui est souvent le cas dans le domaine de la réadaptation auprès de personnes présentant des troubles mentaux. Ainsi, les résultats de ce type d'évaluation permettront d'améliorer les composantes actives des programmes avant de se lancer dans une évaluation des effets.

RÉFÉRENCES

BLISS, M.J. et J.G. EMSHOFF (2002). *Workbook for Designing a Process Evaluation*, Atlanta, Georgia Department of Human Resources, Division of Public Health, Georgia State University.

BRIAND, C., M.-J. DURAND, L. ST-ARNAUD et M. CORBIÈRE (2007). « Work and mental health : Learning from return-to-work rehabilitation programs designed for workers with musculoskeletal disorders », *International Journal of Law and Psychiatry*, vol. 30, n°s 4-5, p. 444-457.

CHAMPAGNE, F., A. BROUSSELLE, Z. HARTZ, A.P. CONTANDRIOPOULOS et J.-L. DENIS (2011). « L'analyse de l'implantation », dans A. Brousselle, F. Champagne, A.P. Contandriopoulos et Z. Hartz (dir.), *L'évaluation : concepts et méthodes*, 2e éd., Montréal, Les Presses de l'Université de Montréal, p. 237-273.

CHEN, H.T. (2005). *Practical Program Evaluation. Assessing and Improving Planning, Implementation, and Effectiveness*, Thousand Oaks, Sage Publications.

CONTANDRIOPOULOS, A.P., F. CHAMPAGNE, J.L. DENIS et M.C. AVARGUES (2000). « L'évaluation dans le domaine de la santé : concepts et méthodes », *Revue d'épidémiologie et de santé publique*, vol. 48, p. 517-539.

DAWIS, R.V. et L.H. LOFQUIST (1984). *A Psychological Theory of Work Adjustment*, Minneapolis, University of Minnesota.

DAWIS, R.V., L.H. LOFQUIST et D.J. WEISS (1968). *A Theory of Work Adjustment (A Revision)*, Minneapolis, University of Minnesota.

DOBSON, D. et T.J. COOK (1980). « Avoiding type III errors in program evaluation : Results from a field experiment », *Evaluation and Program Planning*, vol. 3, nº 4, p. 269-376.

DURAND, M.-J., D. BERTHELETTE, P. LOISEL et D. IMBEAU (2012a). « Étude de la fidélité de l'implantation d'un programme de réadaptation au travail auprès de travailleurs de la construction ayant une dorsolombalgie », *Perspectives interdisciplinaires sur le travail et la santé (PISTES)*, vol. 14, nº 1, <http://www.pistes.uqam.ca/v14n11/articles/v14n11a12.htm>.

DURAND, M.-J., D. BERTHELETTE, P. LOISEL et D. IMBEAU (2012b). « Validation of the programme impact theory for a work rehabilitation programme », *Work : A Journal of Prevention, Assessment & Rehabilitation*, vol. 42, nº 4, p. 495-505.

DURAND, M.-J. et P. LOISEL (2001a). « La transformation de la réadaptation au travail d'une perspective parcellaire à une perspective systémique », *Perspectives interdisciplinaires sur le travail et la santé (PISTES)*, vol. 3, nº 2, <http://www.pistes.uqam.ca/v3n2/pdf/v3n2a5.pdf>.

DURAND, M.-J. et P. LOISEL (2001b). « Therapeutic return to work : Rehabilitation in the workplace », *Work : A Journal of Prevention, Assessment & Rehabilitation*, vol. 17, nº 1, p. 57-63.

DURAND, M.-J., P. LOISEL, N. CHARPENTIER, J. LABELLE et Q.N. HONG (2004). *Le programme de retour thérapeutique au travail (RTT)*, Longueuil, Centre de recherche clinique en réadaptation au travail PRÉVICAP de l'Hôpital Charles-Le Moyne.

DURAND, M.-J., P. LOISEL, Q.N. HONG et N. CHARPENTIER (2002). « Helping clinicians in work disability prevention : The Work Disability Diagnosis Interview », *Journal of Occupational Rehabilitation*, vol. 12, nº 3, p. 191-204.

DURAND, M.-J., B. VACHON, P. LOISEL et D. BERTHELETTE (2003). « Constructing the program impact theory for an evidence-based work rehabilitation program for workers with low back pain », *Work : A Journal of Prevention, Assessment and Rehabilitation*, vol. 21, nº 3, p. 233-242.

ÉQUIPE CREXE (2010). *Évaluation du Programme régional de réadaptation pour les personnes atteintes de maladies chroniques (programme Traité santé)*, Montréal, École national d'administration publique.

FIXSEN, D.L., S.F. NAOOM, K.S. BLASE, R.M. FRIEDMAN et F. WALLACE (2005). *Implementation Research : A Synthesis of the Literature*, Tampa, University of South Florida.

FORTIN, M.-F. (2010). *Fondements et étapes du processus de recherche*, 2e éd., Québec, Chenelière Éducation.

JABOT, F. et M. BAUCHET (2009). « Glossaire », dans V. Ridde et C. Dagenais (dir.), *Approches et pratiques en évaluation de programme*, Montréal, Les Presses de l'Université de Montréal, p. 327-332.

LANDRY, R. (1997). « L'analyse de contenu », dans B. Gauthier (dir.), *Recherche sociale. De la problématique à la collecte de données*, Québec, Presses de l'Université du Québec, p. 329-356.

LETTS, L., M. LAW, N. POLLOCK, D. STEWART, M. WESTMORLAND, A. PHILPOT *et al.* (1999). *L'évaluation de programme en ergothérapie. Guide pratique fondé sur l'évidence scientifique*, Ottawa, Association canadienne des ergothérapeutes.

LINNAN, L. et A. STECKLER (2002). « An overview », dans A. Steckler et L. Linnan (dir.), *Process Evaluation for Public Health Interventions and Research*, San Francisco, John Wiley & Sons, p. 1-23.

LOVE, A. (2004). « Implementation evaluation », dans J.S. Wholey, H.P. Hatry et K.E. Newcomer (dir.), *Handbook of Practical Program Evaluation*, San Francisco, John Wiley & Sons, p. 63-97..

MCLAUGHLIN, J.A. et B.J. GRETCHEN (2010). « Using logic models », dans J.S. Wholey, H.P. Hatry et K.E. Newcomer (dir.), *Handbook of Practical Program Evaluation*, 3e éd., San Francisco, John Wiley & Sons, p. 55-80.

MUHR, T. (2004). *User's Manual for ATLAS.ti*, 5.0, Berlin, ATLAS.ti Scientific Sofware Development.

NADEAU, M.-A. (1988). *L'évaluation de programme. Théorie et pratique*, Québec, Les Presses de l'Université Laval.

OAKLEY, A., V. STRANGE, C. BONELL, E. ALLEN et J. STEPHENSON (2006). « Process evaluation in randomized controlled trials of complex interventions », *British Medical Journal*, vol. 332, no 7538, p. 413-416.

PATTON, M.Q. (2008). *Utilization-Focused Evaluation*, 4e éd., Thousand Oaks, Sage Publications.

PAULS, M. (2004). *Évaluation du Programme de mieux-être communautaire Mamowichihitowin : Étape 1 – description et modèle logique du programme*, Calgary, Institut canadien de la recherche sur le droit de la famille.

PORTEOUS, N. (2009). « La construction du modèle logique d'un programme », dans V. Ridde et C. Dagenais (dir.), *Approches et pratiques en évaluation de programme*, Montréal, Les Presses de l'Université de Montréal, p. 87-105.

PROVUS, M. (1971). *Discrepancy Evaluation for Educational Program Improvement and Assessment*, Berkeley, McCutchan.

ROSSI, P.H., M.W. LIPSEY et H.E. FREEMAN (2004). *Evaluation. A Systematic Approach*, 7e éd., Thousand Oaks, Sage Publications.

SLY, A. et K. TAYLOR (2005). *Évaluation du programme de réadaptation psychosociale en milieu de vie structuré à l'intention des délinquantes*, Ottawa, Direction générale de la recherche, Service correctionnel du Canada.

YIN, R.K. (2003). *Case Study Research : Design and Methods*, 3e éd., Thousand Oaks, Sage Publications.

CHAPITRE 10

L'ESSAI CONTRÔLÉ RANDOMISÉ
Illustration pour déterminer l'efficacité d'une intervention psychosociale

Tania Lecomte

FORCES

- Il est une approche standardisée incluant des critères très précis qui s'avèrent faciles à suivre.
- La randomisation assure une comparaison rigoureuse.
- La préconisation d'évaluateurs «à l'insu» permet une évaluation non biaisée.

LIMITES

- Il ne représente pas la réalité des milieux cliniques.
- Les préférences des participants ne sont pas considérées dans un essai randomisé.
- Ses résultats tiennent rarement compte de la dose thérapeutique nécessaire à l'efficacité de l'intervention évaluée.

L'essai contrôlé randomisé ou aléatoire constitue, dans le domaine de la santé, l'étape nécessaire afin de déterminer l'efficacité d'un traitement, d'une intervention ou d'un programme ou service. Quoique les études randomisées en psychologie existent depuis la fin du XIX[e] siècle, l'essai contrôlé randomisé est devenu un devis de recherche incontournable en santé seulement depuis la fin du XX[e] siècle (Meldrum, 2000). Cette réalité est issue de l'évaluation des interventions psychosociales dans le but de déterminer la supériorité d'une nouvelle intervention (ou d'un nouveau traitement ou programme) lorsque cette dernière est comparée avec le traitement usuel ou avec une intervention de comparaison. L'essai contrôlé randomisé comprend des étapes prédéfinies très strictes (que nous détaillerons plus loin) et permet de répondre à certaines questions, notamment celles qui concernent l'efficacité d'une intervention ou d'un programme dans un contexte donné. Cependant, l'essai contrôlé randomisé ne permet pas de répondre à d'autres questions telles que « Quelle est l'utilité de l'intervention en contexte réel ? » ou encore « Quels sont les facteurs qui expliquent les changements chez l'individu ? ». Dans le but d'illustrer les étapes requises pour un essai contrôlé randomisé, une étude visant à déterminer l'efficacité d'une intervention cognitive comportementale de groupe pour les personnes ayant récemment vécu un premier épisode psychotique sera présentée.

1. DÉFINITION D'UN ESSAI CONTRÔLÉ RANDOMISÉ

Un essai contrôlé randomisé (ECR) est un devis utilisé pour une étude expérimentale dans laquelle les chercheurs répartissent de manière aléatoire les participants en deux ou plusieurs groupes qui recevront ou non l'intervention étudiée, et qui feront par la suite l'objet de comparaisons intergroupes. L'étude expérimentale vise à vérifier une hypothèse, telle que l'efficacité d'un traitement, en manipulant soit le traitement (durée, intensité, etc.), soit qui le reçoit. Parmi les devis utilisés lors d'études expérimentales, le devis ECR est le plus répandu. Ce devis est essentiel pour déterminer l'efficacité d'un traitement médical, comme l'évaluation d'une nouvelle molécule pharmaceutique. Il est intéressant de noter que l'ECR est maintenant considéré comme une étape essentielle pour évaluer l'efficacité de la plupart des nouvelles interventions psychosociales telles que les psychothérapies. L'avantage premier de l'ECR est qu'il permet de montrer rigoureusement qu'une intervention offre des bénéfices pour l'individu lorsque celle-ci est comparée à une intervention contrôle ou à un traitement usuel, et ce, peu importe le type d'individu qui reçoit l'intervention, pour autant qu'il réponde aux critères d'inclusion. L'ECR assure de par son devis que les

participants ne sont pas sélectionnés *a priori* pour recevoir ou non une intervention, et donc que les résultats sont liés à l'intervention et non au profil particulier des participants. Un ECR rigoureux suit une série de règles comme l'utilisation d'évaluateurs qui ne doivent pas être influencés par les effets attendus de l'étude ou encore une intervention contrôle permettant de comparer les effets essentiels de l'intervention. La plupart des ECR visent à déterminer la supériorité d'une intervention sur une autre. Il est toutefois possible d'effectuer un ECR visant à vérifier l'équivalence d'une intervention à l'égard d'une autre, ou encore la *non-infériorité* de la nouvelle intervention testée[1].

1.1. Différence entre un ECR médical et un ECR psychosocial

La différence principale entre un devis ECR médical et un devis ECR psychosocial est relative à l'information que le participant reçoit concernant son traitement ou l'intervention qu'il recevra. En effet, dans un essai pharmaceutique, le participant pourra recevoir une pilule et ne saura pas ce qu'elle contient (peut-être une pilule placebo). Pour l'ECR psychosocial, le participant est au courant qu'il reçoit une thérapie ou une intervention et de la nature de celle-ci puisque son implication est nécessaire. En ce sens, il est pratiquement impossible d'offrir des thérapies « placebos » à des individus. Il est à noter cependant que certains auteurs tentent d'offrir des interventions contrôles comparables, mais qui ne possèdent pas les éléments jugés les plus efficaces, espérant ainsi offrir une forme de traitement placebo. Un bon exemple de cette approche est l'étude de Sensky et collègues (2000), qui ont offert au groupe témoin des sessions de discussions amicales à titre de comparaison avec une thérapie cognitive comportementale à l'étude. Le groupe témoin s'est avéré être quasiment aussi efficace que l'intervention supposée « placebo », car celle-ci comportait plusieurs éléments thérapeutiques essentiels, notamment une interaction sympathique avec une personne attentive et intéressée au participant et à son bien-être. Il est important de noter que l'essai contrôlé randomisé utilise seulement des critères objectifs lors du recrutement (p. ex., présence du diagnostic X et de symptômes Y), ce qui fait en sorte qu'il est souvent obligatoire de faire par la suite une étude plus écologique, c'est-à-dire dans le milieu, *in situ*, avec des intervenants du milieu. Ainsi, il sera possible de déterminer si l'intervention jugée efficace peut aussi être efficiente, c'est-à-dire offrir des résultats

1. L'équivalence et la non-infériorité veulent essentiellement dire la même chose, mais certains chercheurs tentent parfois de démontrer qu'une intervention d'apparence moins efficace (p. ex., une thérapie par Internet) peut être suffisamment efficace pour être jugée « non inférieure » à l'intervention de prédilection dans la littérature scientifique (p. ex., la thérapie en face à face).

significatifs en contexte naturel. En effet, les intervenants vont utiliser des critères différents, comme la motivation de l'individu ou encore ses capacités relationnelles ou cognitives, avant de déterminer qui sera susceptible de recevoir une intervention.

1.2. Règles à suivre et règles de présentation CONSORT

Le CONSORT est une liste détaillée d'informations qui doit être incluse dans l'article présentant les résultats de l'essai contrôlé randomisé (Schulz, Altman et Moher, 2010). Beaucoup l'utilisent comme un guide pour s'assurer que tous les éléments nécessaires à l'ECR ont été respectés. Bien que le CONSORT soit un point de référence utile, il ne contient pas toutes les informations requises à la conduite d'un essai contrôlé randomisé. À cet effet, le CONSORT est générique, conçu pour des études à un niveau international, alors que certains pays exigent des règles particulières pour la tenue d'un essai contrôlé randomisé, telles que l'enregistrement de l'ECR sur un site Internet en Amérique du Nord (<http://www.clinicaltrials.gov>). Mis à part la liste détaillée d'éléments à inclure, le CONSORT a aussi l'avantage de proposer un tableau décrivant la distribution dans le temps des participants à l'étude (voir la figure 10.1). D'ailleurs, ce tableau descriptif est maintenant devenu une exigence pour la plupart des revues scientifiques. Il existe aussi un instrument de mesure développé par Tarrier et Wykes (2004), nommé le CTAM (Clinical Trials Assessment Measure). Cette mesure consiste en des questions dont les réponses permettent de déterminer le niveau de qualité d'un ECR portant sur une intervention psychosociale. Les réponses pour chacune des questions sont quantifiées et un score total de 100 % indique que les règles de l'art pour mener un ECR ont toutes été respectées (voir le tableau 10.2). Essentiellement, les grandes règles à suivre peuvent être divisées en huit éléments, qui vont être décrits ci-après et résumées dans le tableau 10.1.

1.2.1. Structure de l'ECR et choix de l'intervention de comparaison

Avant de commencer l'ECR pour évaluer une intervention psychosociale, il est essentiel de vérifier si ce type d'ECR a déjà été réalisé dans d'autres études ou contextes. Cette information peut être obtenue, d'une part, en réalisant une recension des écrits et, d'autre part, en consultant le site <http://www.clinicaltrials.gov> pour vérifier si une étude similaire y a été enregistrée. Aussi, il est possible d'obtenir cette information lors de rencontres informelles avec des collègues du même champ disciplinaire lors de congrès scientifiques. Dans le cas où ce type d'étude a été réalisé, il pourrait être

important de déterminer quelles sont les recommandations qui ont été proposées par les auteurs en vue d'apporter d'éventuelles améliorations. Il est aussi suggéré d'utiliser des outils de mesure similaires dans le but de comparer les résultats d'un ECR à un autre. Dans le cas où l'étude est entièrement innovatrice, il demeure crucial et fortement recommandé d'utiliser des outils de mesure valides et standardisés qui existent en anglais (ou mieux encore, traduits en plusieurs langues). Cette précaution permettra à d'autres chercheurs de répliquer au besoin l'ECR. Il est aussi important de bien choisir l'intervention de comparaison en considérant plusieurs critères. Cette dernière devrait idéalement avoir la même structure et durée que l'intervention psychosociale évaluée et comprendre un volume d'interactions similaire (p. ex., une intervention de groupe est comparée avec une autre intervention de groupe de même durée). Si de surcroît l'intervention de comparaison est une intervention thérapeutique jugée efficace, alors l'étude sera considérée comme plus rigoureuse. Dans ce dernier cas, il faudra cependant envisager une taille d'échantillon plus conséquente afin de repérer des différences significatives, surtout si l'on cherche à déterminer la supériorité d'une intervention sur une autre. À titre indicatif, une alternative intéressante est d'utiliser une intervention contrôle efficace, mais qui n'a pas exactement les mêmes cibles thérapeutiques. Ainsi, les cibles précises de chaque intervention peuvent être évaluées singulièrement. On peut penser par exemple à une intervention de groupe visant l'amélioration de la mémoire comme intervention contrôle dans le contexte d'un ECR évaluant une intervention de groupe visant l'affirmation de soi.

La plupart des ECR utilisent des groupes parallèles égaux (c.-à-d. une randomisation égale de 1 pour 1). S'il s'agit d'une étude répliquée et que l'on sait *a priori* que l'intervention contrôle (de comparaison) offrira moins d'effets significativement positifs que la nouvelle intervention proposée, il est alors possible de faire une étude randomisée à groupes inégaux. On peut ainsi retrouver un ratio de 2 pour le groupe expérimental contre 1 pour le groupe témoin, ou encore faire en sorte que les participants du groupe témoin puissent avoir accès à l'intervention après une période d'attente (liste d'attente). Toutefois, les analyses statistiques seront alors plus complexes à effectuer.

1.2.2. Vérification des bienfaits de l'étude

Il est recommandé qu'une équipe de recherche, plutôt qu'un seul et unique chercheur, soit responsable de l'ECR, surtout si l'ECR est implanté dans plusieurs sites. Cette équipe est appelée « comité aviseur » et comprend la plupart du temps les chercheurs de l'étude, le coordonnateur de recherche, un consultant statistique, une personne experte en éthique, un clinicien et

un représentant de la clientèle visée par l'étude. Parfois, une même personne peut jouer plusieurs rôles, mais la présence d'un comité aviseur est essentielle pour offrir un éventail varié de points de vue, tout en respectant les règles d'un ECR et le bien-être des participants. Les membres du comité aviseur se rencontrent de manière périodique afin de discuter de l'évolution de l'étude et de s'ajuster aux divers problèmes pouvant survenir, et aussi pour vérifier au besoin si l'étude doit se poursuivre advenant l'apparition d'effets néfastes. Même si en principe un essai contrôlé randomisé exige que les données ne soient pas analysées avant la fin de l'étude, s'il s'agit d'une nouvelle intervention ou programme, il est recommandé de « briser l'insu » (*break the blind*) une seule fois, vers le début de l'étude, lorsque suffisamment de données ont été recueillies (p. ex., une vingtaine de participants), et ce, dans le but de déterminer si l'intervention ciblée semble être thérapeutique pour les participants ou si, au contraire, elle a des effets néfastes pour ces derniers. À la suite de cette vérification, les données recueillies ne seront pas analysées avant la fin de la collecte des données.

1.2.3. Description de la clientèle visée

Les critères d'inclusion doivent être clairs, détaillés et représenter autant que possible la clientèle visée. Il n'est pas recommandé d'être trop exclusif dans le choix des critères d'acceptation des participants, afin de représenter la clientèle ciblée dans l'ensemble de ses caractéristiques. Par ailleurs, dans le cadre d'un ECR, les réviseurs d'articles s'attendent à une description détaillée de l'échantillon, notamment grâce aux diagnostics psychiatriques obtenus lors d'entrevues semi-structurées plutôt que par la prise de notes dans les dossiers médicaux des participants, qui ne s'avèrent pas assez rigoureuses (p. ex., l'évaluation par l'entrevue SCID [Structured Clinical Interview for DSM-IV] ; First *et al.*, 1997). Les critères d'exclusion doivent être non seulement clairs et précis, mais aussi bien justifiés.

1.2.4. Qualité et fidélité de l'intervention

L'intervention doit être standardisée, c'est-à-dire qu'elle doit être décrite dans un manuel avec suffisamment de détails pour que d'autres intervenants puissent (à la suite d'une formation ou non) l'utiliser de manière fiable. Dans le cadre d'un ECR, il est souvent nécessaire d'enregistrer (audio ou vidéo) les sessions où l'intervention est offerte afin de vérifier si les intervenants suivent le manuel à la lettre et offrent l'intervention telle qu'elle a été conçue. Des juges indépendants visionnent ou écoutent alors une portion importante des enregistrements (minimum 50 % pour chaque intervention, incluant l'intervention de comparaison) en cotant

de manière indépendante les entrevues et en utilisant une grille de fidélité. Il est fortement recommandé d'utiliser des grilles précises permettant de vérifier la présence de certaines composantes reconnues comme essentielles à l'intervention. Il n'est pas suffisant, comme cela a été réalisé dans certaines études, de demander à des assistants de recherche de tenter de deviner l'approche thérapeutique à la suite de l'écoute des enregistrements. Cette méthode intuitive présente plusieurs écueils, comme on peut s'en douter. Une évaluation plus pointue portant sur les techniques utilisées est donc préférable.

Les cotations indépendantes obtenues permettront de déterminer le degré de fidélité interjuges (c.-à-d. s'ils sont d'accord ou non sur leurs cotes) et ainsi de refléter la qualité de la fidélité à l'intervention. Si la fidélité interjuges est jugée trop basse, un autre évaluateur, voire un des chercheurs de l'équipe, devra visionner les sessions et les cotes problématiques afin de décider quel évaluateur devra modifier sa façon de coter.

1.2.5. Choix de la résultante et de la taille de l'effet

Il est recommandé d'avoir une mesure de résultante principale de l'étude (liée aux effets attendus de l'intervention) et quelques mesures de résultats considérées comme secondaires. Le calcul de puissance et la taille de l'effet souhaitée sont calculés seulement sur la résultante principale. Suggérer un trop grand nombre de mesures principales exigerait une taille d'échantillon trop grande. Lorsque la taille de l'effet est calculée, il est nécessaire de tenir compte de l'effet attendu. Toutefois, si une petite taille de l'effet est visée, alors il est attendu de recruter un échantillon plus conséquent, ce qui entraîne en l'occurrence de conduire l'étude sur plusieurs années. Il est important de se fier aux études déjà publiées dans le domaine et, en général, une taille de l'effet moyenne est souvent une bonne estimation. Cependant, il est de plus en plus courant de lire des études publiées n'ayant pas suffisamment de puissance pour pouvoir détecter des effets significatifs – les conclusions ne permettent donc pas de se prononcer sur l'efficacité de l'intervention alors qu'un échantillon plus grand aurait peut-être révélé une différence significative. De telles études ne sont pas très utiles ni recommandées, même si une métaanalyse subséquente incluant plusieurs de ces études pourrait offrir des conclusions intéressantes. En bref, nous n'insisterons jamais assez pour dire que la taille de l'échantillon est un critère incontournable pour pouvoir tirer des conclusions claires. En ce sens, prendre connaissance des études déjà publiées en tirant profit des obstacles rencontrés par les auteurs lors de la réalisation de leur étude reste une méthode judicieuse.

1.2.6. Sélection aléatoire

La sélection aléatoire devrait être basée sur la chance et ne pas être prévisible. Plusieurs méthodes de randomisation existent ; parmi elles, on note : l'utilisation d'une matrice ou d'un tableau générant une séquence de chiffres liés aux participants, ou encore un programme informatique permettant de randomiser les sujets selon un ou des critères prédéterminés (p. ex., <http://www.randomization.com>). Parmi les critères, il est possible de retenir une distribution égale entre les groupes en tenant compte de l'âge, du genre, ou encore d'une cote de base sur une mesure en particulier (p. ex., fonctionnement social). L'utilisation du « tirage dans un chapeau » n'est plus considérée comme une méthode valide de nos jours. Il est aussi important de prendre en note le nombre de personnes approchées pour participer à l'étude, le nombre de personnes ayant refusé l'intervention ou l'étude, et dans la mesure du possible les raisons sous-jacentes à ces refus, ainsi que le nombre et les raisons d'abandon pendant l'ECR. Toutes ces informations devront être incluses dans le diagramme CONSORT lors de la publication des résultats.

1.2.7. Évaluation à l'insu

L'ECR doit être effectué au moins en partie à l'insu, ou comporter un « masque », c'est-à-dire que les personnes engagées dans la collecte des données (p. ex., les assistants de recherche) ne doivent pas connaître ou ne doivent pas être influencées par le fait qu'un participant a reçu une intervention ou une autre. Certaines stratégies peuvent être utilisées pour assurer cette méthode à l'aveugle. Premièrement, il est suggéré que les évaluateurs soient indépendants des chercheurs ou des thérapeutes participant à l'étude. Deuxièmement, certaines mesures ou tests préintervention peuvent être administrés avant que les participants ne soient randomisés. Troisièmement, les participants peuvent être informés de ne rien révéler à l'évaluateur en ce qui concerne l'intervention qu'ils ont reçue. Dans le cas où il est difficile pour les participants de ne rien révéler, une nouvelle stratégie est alors mise en œuvre. Il s'agit d'avoir plusieurs évaluateurs pour l'ECR tout en s'assurant que ces derniers présentent une fidélité interjuges satisfaisante, et d'alterner les évaluateurs selon le temps de mesure (c.-à-d. que le même évaluateur n'évalue pas le même participant deux fois de suite). Ainsi, l'évaluateur ne peut pas se « souvenir » d'une cote antérieure donnée au participant, ce qui limite la probabilité qu'il perçoive le participant d'un œil différent.

1.2.8. Type d'analyses statistiques

Les analyses statistiques utilisées doivent tenir compte de plusieurs facteurs. La plupart des auteurs recommandent des analyses « intention de traitement » (*intent to treat*) (Tarrier et Wykes, 2004). Ces analyses permettent d'inclure les données d'un participant qu'il ait ou non complété le traitement, sous prétexte qu'une faible dose de l'intervention serait suffisante pour obtenir des résultats positifs et significatifs. Par exemple, lorsqu'il s'agit d'une molécule comme une pilule pour enlever les maux de tête, il est possible qu'une seule dose suffise pour constater une résultat. Cependant, en ce qui concerne la plupart des interventions psychosociales, la relation dyadique (alliance thérapeutique) entre l'intervenant et son client est requise avant de pouvoir bénéficier pleinement de l'approche en tant que telle, ce qui engendre le besoin d'avoir plusieurs sessions afin d'établir une bonne alliance thérapeutique. Ainsi, la plupart des chercheurs préfèrent atteindre une certaine « dose » de l'intervention (p. ex., un minimum de 30 % ou 50 % de toutes les sessions) pour ainsi s'assurer que les résultats sont inhérents aux composantes de l'intervention. D'autres chercheurs vont seulement inclure dans leurs statistiques les personnes ayant participé à l'intervention dans sa totalité, cherchant ainsi à déterminer l'effet optimal de l'intervention.

Souvent les statistiques rapportées pour un essai à trois groupes (p. ex., une nouvelle intervention, une intervention de comparaison et un groupe d'attente ou de traitement régulier) avec un résultat principal de type continu ou catégoriel sont des Manova (voir le chapitre 16 de cet ouvrage), mais de plus en plus on retrouve des analyses plus sophistiquées permettant d'appréhender des modèles hiérarchiques linéaires (voir le chapitre 20 de cet ouvrage) pour tenir compte de l'effet de cohorte dans le temps (les participants ne sont pas tous recrutés en même temps et dans le même lieu et n'ont pas tous les mêmes thérapeutes). Quoique l'information sur l'effet statistiquement significatif de l'intervention soit importante, il est aussi nécessaire de chercher à savoir si l'effet est aussi cliniquement significatif et de présenter cet élément. Pour déterminer un effet cliniquement significatif, il importe de vérifier les valeurs de référence de l'instrument qui a été utilisé et de se poser, en termes cliniques, la question suivante : une diminution de symptômes de deux points est-elle considérée comme significative selon les valeurs de référence ou les auteurs de l'instrument ?

Tableau 10.1.
Règles pour un essai contrôlé randomisé

1. Structure de l'essai et choix du contrôle : Vérifier si l'étude a déjà été faite. S'inscrire sur un site répertoriant les essais contrôlés randomisés. Choisir une intervention contrôle appropriée (idéalement avec des vertus thérapeutiques).
2. Vérification des bienfaits de l'étude : Mettre sur pied un comité aviseur. Vérifier s'il y a des effets néfastes une seule fois, en début d'étude.
3. Description de la clientèle visée : Avoir des critères d'inclusion clairs, précis, et représentant autant que possible la clientèle visée.
4. Qualité et fidélité de l'intervention : Utiliser une intervention standardisée, idéalement manualisée et vérifier si l'intervention est effectivement offerte selon les standards demandés.
5. Choix d'une résultante principale et de la taille de l'effet : Choisir une seule résultante principale et quelques mesures secondaires. Faire le calcul de puissance pour déterminer la taille de l'échantillon sur la mesure principale.
6. Sélection aléatoire : S'assurer que la sélection est réellement aléatoire. Noter les raisons de refus et d'abandons pour construire le diagramme CONSORT lors de la publication des données.
7. Évaluation à l'insu : Mettre en œuvre de bonnes stratégies à l'insu, pour assurer l'impartialité des intervieweurs ou des assistants de recherche lors des évaluations.
8. Type d'analyses statistiques : Choisir les analyses statistiques appropriées selon le plan et l'hypothèse principale de l'étude.

2. ILLUSTRATION : ESSAI CONTRÔLÉ RANDOMISÉ D'UNE THÉRAPIE COGNITIVE COMPORTEMENTALE DE GROUPE POUR PERSONNES EN DÉBUT DE PSYCHOSE

Dans cette partie, l'essai contrôlé randomisé de la thérapie cognitive comportementale de groupe offerte à des personnes en début de psychose sera illustré en considérant les huit règles à suivre qui ont été présentées en première partie de ce chapitre.

2.1. Structure de l'essai et choix du contrôle

Lors de la recension des écrits conduite sur le sujet, une métaanalyse couvrant plusieurs études a montré que la thérapie cognitive comportementale (TCC) permettait de diminuer les symptômes psychotiques chez les personnes ayant une expérience prolongée de la psychose (Gould *et al.*, 2001). Cependant, aucune étude ne s'était intéressée à l'effet de cette intervention sur une clientèle en début de psychose. Or les croyances irrationnelles (ou délirantes) des personnes en début de psychose sont souvent moins rigides et donc plus susceptibles d'être modifiées par une intervention cognitive, contrairement à une croyance existant depuis plusieurs années. De plus, la TCC sous le format de groupe plutôt qu'en individuel (c.-à-d. en face à face) semblait un aspect novateur. En effet, l'intervention

de groupe est une modalité assez courante en Amérique du Nord et semblait plus facile à implanter qu'une modalité individuelle, vu le nombre réduit de psychologues en milieu psychiatrique. Par ailleurs, notre objectif était de choisir deux groupes témoins : une intervention visant une résultante similaire, mais en proposant des techniques différentes, et un groupe « liste d'attente ». L'intervention contrôle retenue était l'entraînement aux habiletés sociales visant la gestion des symptômes. L'entraînement aux habiletés sociales est une modalité d'intervention reconnue comme fondée sur des résultats probants en Amérique du Nord. Pourtant, aucune étude à notre connaissance n'avait tenté de comparer les effets de la TCC à celle de l'entraînement à la gestion des symptômes (une des formes d'entraînement aux habiletés sociales). L'étude proposée visait donc à comparer une intervention TCC de groupe pour les personnes en début de psychose avec une intervention de groupe pour la même population qui porte sur l'entraînement à la gestion des symptômes, ainsi qu'à un groupe « liste d'attente » présentant les mêmes caractéristiques. Enfin, une fois les fonds obtenus des Instituts de recherche en santé du Canada pour réaliser cette étude, nous avons enregistré notre ECR sur le site <http://www.clinicaltrials.org>.

2.2. Vérification des bienfaits de l'étude

Dans le cadre de cette étude, un comité aviseur officiel n'a pas été mis en place. Toutefois, une équipe composée des directeurs des cliniques de premier épisode, de trois chercheurs et de la coordonnatrice de recherche échangeait régulièrement, notamment à propos des obstacles liés au bon déroulement de l'étude (voir Spidel, Lecomte et Leclerc, 2006). Les résultats de la première cohorte ont été analysés pour vérifier que l'intervention se déroulait correctement et sans effets néfastes (Lecomte *et al.*, 2003).

2.3. Description de la clientèle visée

Les critères d'inclusion de l'étude étaient les suivants : 1) toute personne en début de psychose ; 2) suivie depuis moins de deux ans dans une clinique de premier épisode ; 3) âgée entre 18 et 35 ans ; 4) présentant des symptômes psychotiques stables ou récurrents lors de périodes de stress ; et 5) sachant communiquer (lire et écrire) en français ou en anglais. Les participants ne devaient pas recevoir d'autres interventions à l'exception du traitement régulier offert par la clinique. Les changements de médication étaient mesurés pour vérifier leur lien avec la résultante. Une entrevue diagnostique validée était utilisée, soit le SCID (First *et al.*, 1997), afin de bien décrire les diagnostics psychiatriques des participants à l'étude. Les critères

d'exclusion étaient liés à des difficultés de lecture ou d'écriture (chaque participant devait lire et écrire dans son manuel d'intervention), ou encore à des déficits cognitifs graves (QI de moins de 70). Par ailleurs, les personnes présentant un problème de consommation de substances psychoactives ou d'alcool, c'est-à-dire ayant un trouble concomitant, pouvaient participer à l'étude, car ce trouble est très fréquent chez la clientèle visée (environ 50 %) (Mueser *et al.*, 2001).

2.4. Qualité et fidélité de l'intervention

Les interventions étaient manualisées et standardisées. L'intervention de groupe TCC comprenait un module conçu par notre équipe en consultation avec plusieurs cliniciens, et consistait en 24 rencontres d'une heure (Lecomte *et al.*, 2003). Le module était divisé en quatre parties : 1) le stress ; 2) les techniques TCC ; 3) l'alcool, la drogue et les émotions difficiles ; et 4) les stratégies d'adaptation (*coping*). L'intervention de contrôle portant sur la gestion des symptômes (entraînement aux habiletés sociales) comprenait aussi un module de 24 rencontres d'une heure, dont les sessions étaient : 1) reconnaître les signes de rechute ; 2) gérer les signes de rechute ; 3) reconnaître les symptômes persistants ; et 4) refuser l'alcool et les drogues. Chaque intervention était offerte par deux intervenants ayant chacun reçu une formation de deux jours. Un des intervenants pour chaque groupe provenait de l'équipe de recherche alors que l'autre provenait directement du milieu clinique où les participants étaient recrutés.

Toutes les interventions étaient filmées, tant aux fins de supervision que pour déterminer la fidélité au protocole de traitement selon des échelles validées. La chercheuse principale visionnait les enregistrements afin d'offrir la supervision nécessaire, alors que deux assistants de recherche formés visionnaient 50 % des enregistrements et cotaient la fidélité au protocole de traitement, de manière indépendante et pour les mêmes sessions, soit avec l'échelle CBT Fidelity Scale (Young et Beck, 1980), soit avec le Social Skills Training Fidelity Checklist (Wallace, 2001) selon le groupe évalué. La fidélité interjuges ainsi que les scores obtenus pour chaque sous-échelle des mesures de fidélité ont ensuite été calculés et étaient jugés excellents (coefficient de corrélation intraclasse [CCI] = 0,97-0,98 pour chaque intervention) (Lecomte *et al.*, 2008). Une des difficultés avec les enregistrements était que les participants acceptaient que les thérapeutes soient filmés pour vérifier leur fidélité à l'intervention, mais ne souhaitaient pas être filmés eux-mêmes. Ainsi, on ne voit que l'arrière des têtes des participants dans plusieurs groupes filmés, ce qui rend la supervision plus difficile puisque n'ayant pas accès au non-verbal de plusieurs d'entre eux.

2.5. Choix de la résultante et taille de l'effet

La résultante principale de notre étude était la mesure de symptômes positifs du Brief Psychiatric Rating Scale–Expanded Version (BPRS) (Ventura *et al.*, 1993), une mesure standard de symptômes psychiatriques très utilisée en recherche et en clinique. Puisque les études portant sur la TCC pour la psychose s'étaient jusqu'alors surtout intéressées aux symptômes comme résultante principale, dans le cadre de notre étude, nous avons procédé avec le même choix de concept et de mesure. De plus, grâce à ces études de référence, il a été possible de faire un calcul de puissance sur la base des résultats antérieurs dont l'estimation était de l'ordre de 30 sujets par intervention pour obtenir une taille de l'effet moyenne. D'ailleurs, comme mentionné dans le tableau 10.2 du CTAM, un minimum de 27 sujets par intervention est généralement recommandé. Comme mesures secondaires, deux variables liées aux symptômes (symptômes négatifs et globaux) ainsi que trois mesures psychosociales, soit l'estime de soi (Lecomte, Corbière et Laisné, 2006), les stratégies d'adaptation (Edwards et Baglioni, 1993) et le soutien social (Cutrona et Russel, 1987) ont été retenues. Ces mesures ont été choisies, car elles évaluent des concepts traités dans le cadre de la thérapie. En effet, plusieurs sessions portent sur l'estime de soi et sur les stratégies d'adaptation, alors que le soutien social est un aspect important de la thérapie de groupe.

Tableau 10.2.
La mesure de l'évaluation des essais cliniques (CTAM)

Échantillon – deux questions : valeur maximale = 10	
Q1 :	Est-ce que l'échantillon est un échantillon de commodité (valeur 2) ou une cohorte géographique (valeur 5) ou un échantillon très sélectif, p. ex. : des bénévoles (valeur 0) ?
	Échantillon de commodité – par exemple, des personnes fréquentant une clinique, des patients référés ou une cohorte géographique – tous les patients admissibles dans une zone particulière.
Q2 :	Est-ce que la taille de l'échantillon est supérieure à 27 participants dans chaque groupe de traitement (valeur 5) ou basée sur les calculs de puissance adéquats (valeur 5) ?
	Répartition – trois questions : valeur maximale = 16
Q3 :	Y a-t-il une assignation aléatoire ou une minimisation de l'assignation aux groupes de traitements (si oui, la valeur est de 10) ?
Q4 :	Est-ce que la procédure d'assignation aléatoire est décrite (valeur 3) ?
Q5 :	Est-ce que la procédure d'assignation aléatoire est faite de façon indépendante des essais par l'équipe de recherche (valeur 3) ?
	Évaluation (pour le résultat principal) – cinq questions : valeur maximale = 32
Q6 :	Est-ce que les évaluations sont réalisées par des évaluateurs indépendants et non des thérapeutes (valeur 10) ?

(*suite*)

Tableau 10.2.
La mesure de l'évaluation des essais cliniques (CTAM) (*suite*)

Échantillon – deux questions : valeur maximale = 10	
Q7 :	Est-ce que des évaluations standardisées sont utilisées pour mesurer les symptômes de façon standard (valeur 6) ou est-ce des évaluations idiosyncratiques des symptômes (valeur 3) ?
Q8 :	Est-ce que les personnes qui font les évaluations sont ignorantes face aux groupes d'appartenance ou d'assignation (valeur 10) ?
Q9 :	Est-ce que les méthodes à l'aveugle destinées aux évaluateurs sont décrites de façon adéquate (valeur 3) ?
Q10 :	Y a-t-il vérification que les évaluateurs sont vraiment ignorants de l'assignation des participants (valeur 3) ?
	Groupes témoins – une question : valeur maximale = 16
Q11 :	TAU est un groupe témoin (valeur 6) et/ou un groupe témoin qui tient compte des effets non spécifiques ou des autres traitements établis ou crédibles (valeur 10) ?
	Analyse – deux questions : valeur maximale = 15
Q12 :	Les analyses sont-elles convenables pour le modèle de conception (ou le plan) et le type de mesures résultantes (valeur 5) ?
Q13 :	Les analyses incluent-elles tous les participants de façon aléatoire (on nomme parfois ces analyses « analyses de l'intention de traitement ») (valeur 6) et une enquête adéquate sur les raisons des abandons aux évaluations si le taux d'attrition excède 15 % (valeur 4) ?
	Traitement actifs – trois questions : valeur maximaleum = 11
Q14 :	Est-ce que le traitement était décrit correctement (valeur 3) et est-ce qu'il y avait utilisation d'un protocole de traitement ou d'un manuel (valeur 3) ?
Q15 :	Est-ce que le taux d'observance au protocole de traitement ou la qualité du traitement a été évalué (valeur 5) ?
▪ Lorsque le critère n'est pas rencontré pour une question = 0	
▪ **Total : valeur maximale = 100**	

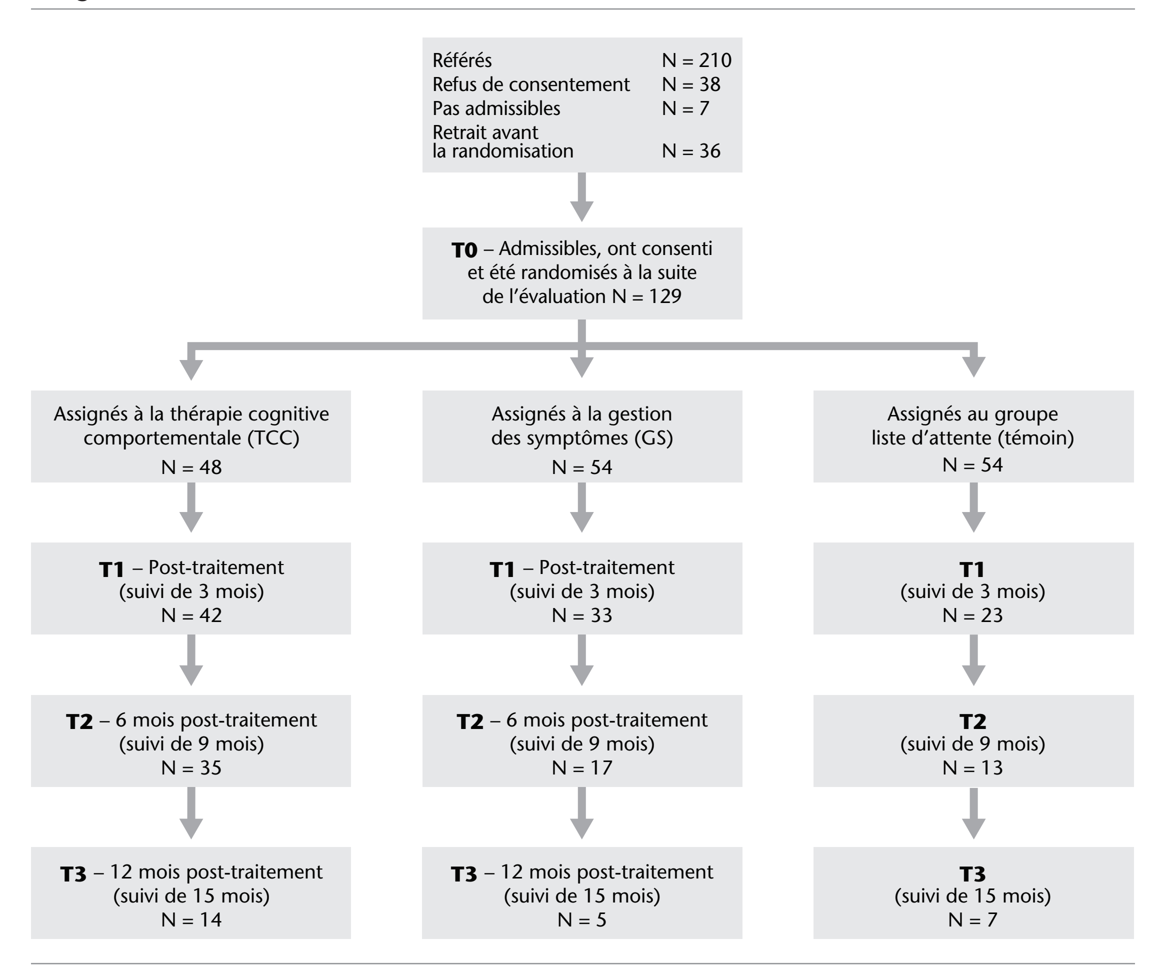
Source : Tarrier et Wykes, 2004, p. 1393 ; traduction libre.

2.6. Sélection aléatoire

La sélection aléatoire a été effectuée par « groupe-cohorte », c'est-à-dire que lorsque suffisamment de personnes étaient recrutées pour constituer un groupe, le groupe était alors randomisé pour recevoir une des deux interventions ou encore mis sur la liste d'attente. Cette forme de randomisation n'est acceptable que dans le contexte d'études portant sur des groupes d'individus et pour lequel une randomisation individuelle impliquerait une attente trop longue avant d'offrir l'intervention, entraînant un risque élevé de perdre les participants avant même que l'intervention débute. Sachant que l'étude comprenait plusieurs sites, le logiciel informatique de randomisation utilisé permettait de tenir compte du milieu (rural, banlieue ou urbain) lors de la randomisation. En vue d'illustrer le nombre

de personnes approchées et ayant terminé l'étude, le diagramme CONSORT est présenté (figure 10.1; pour plus de détails, voir aussi Lecomte, Leclerc et Wykes, 2012). À la lecture des résultats, on peut constater qu'une des deux interventions semblait moins populaire que l'autre et présentait plus d'abandons. Pour compenser cette différence entre les deux interventions, plus de participants ont été recrutés pour cette intervention contrôle. Il est à noter toutefois que cette procédure n'est pas nécessaire pour certaines analyses statistiques dans lesquelles on attribue des scores pour les données manquantes, mais qu'elle devient nécessaire si l'on souhaite observer l'effet de la dose de l'intervention.

Figure 10.1.
Diagramme CONSORT



2.7. Évaluation à l'insu

Pour éviter que les assistants de recherche ne soient influencés par l'information concernant la participation ou non d'une personne à une des deux interventions, trois stratégies ont été utilisées. La première était que la randomisation ne soit réalisée qu'après la première évaluation. À cette étape, les participants eux-mêmes ne savaient pas à quel groupe ils allaient être assignés. La deuxième était que les assistants de recherche ne puissent pas revoir les mêmes participants deux fois de suite, et ainsi ne pouvaient pas avoir de biais de mémoire. Cette stratégie nécessitait que les assistants aient une très bonne fidélité interjuges sur les entrevues utilisées, ce qui était d'ailleurs le cas pour l'étude. Troisièmement, il était demandé aux participants à l'étude de ne pas parler de leur expérience de groupe (qu'elle soit positive, négative ou neutre) avec les assistants de recherche. Enfin, toutes les questions directement liées à l'intervention (p. ex., niveau de satisfaction à l'égard de l'intervention) étaient posées lors de la dernière session de groupe et non lors de la rencontre avec l'assistant de recherche.

2.8. Type d'analyses statistiques

Les analyses préconisées dans cette étude devaient tenir compte de l'effet de cohorte, car les participants provenaient de différents sites cliniques et de milieux de vie (p. ex., urbain, banlieue, rural) et, qui plus est, avaient reçu l'intervention par des thérapeutes différents, ne possédant pas tous les mêmes qualités thérapeutiques. En conséquence, le recrutement a pris quelques années avant d'atteindre la taille d'échantillon escomptée. Pour effectuer les analyses et pour respecter l'effet de cohorte, un modèle hiérarchique linéaire (HLM) incluant l'intention d'être traité a été préconisé. Ainsi, le sujet est niché dans une cohorte, laquelle est nichée dans une condition d'intervention (TCC pour la psychose, gestion des symptômes par entraînement des habiletés sociales [GS] ou liste d'attente). Pour notre étude, cette analyse a démontré que la TCC de groupe était efficace pour diminuer les symptômes généraux (–6,95, $p < 0,05$, taille de l'effet = 0,54), alors que les deux interventions étaient efficaces pour diminuer les symptômes psychotiques (–0,29, –0,48, $p < 0,05$) à la suite de la thérapie. Seule la TCC de groupe offrait des résultats significatifs sur des variables psychosociales, notamment l'augmentation de l'estime de soi (10,32, $p < 0,05$) et des stratégies d'adaptation au stress (8,77, $p < 0,05$). Aussi, une tendance pour une augmentation du soutien social perçu était observée (au suivi de six mois : 6,93, $p = 0,07$). Dans un contexte d'étude différent, où les participants proviendraient tous du même milieu clinique, une Ancova à mesures répétées avec comme covariable la mesure initiale prétraitement

(contrôlant ainsi les différences intergroupes de départ) aurait été suffisante (voir exemple dans Lecomte *et al.*, 1999). Enfin, les effets de nature clinique reconnus comme significatifs ont été évalués aux différents temps de mesure, en tenant compte des valeurs de référence établies pour la mesure de symptôme utilisée, soit le BPRS (Lukoff, Nuechterlein et Ventura, 1986) et des critères du Reliable Change Index de Jacobson et Truax (1991). En ce sens, dans notre étude, les améliorations portant sur les symptômes et l'estime de soi étaient cliniquement significatifs.

CONCLUSION

Bien que l'essai contrôlé randomisé soit encore considéré par plusieurs comme une étape essentielle dans la conception et la validation d'une nouvelle intervention ou d'un nouveau service, elle comporte toutefois certaines limites. L'ECR est implanté notamment dans un contexte souvent artificiel et différent de la réalité clinique habituelle, et le participant ne peut pas choisir l'intervention qu'il recevra. De plus, il est possible que le profil des personnes acceptant de participer à un essai contrôlé randomisé, impliquant souvent plusieurs heures d'évaluations et de tests, soit différent du profil des personnes recevant normalement des services cliniques. L'ECR permet généralement de déterminer si l'intervention est efficace, mais ne permet pas de déterminer à quoi sont dus les effets ou qui en bénéficie le plus. Des études de processus sont souvent nécessaires pour répondre à ce type de questions. En ce sens, plusieurs chercheurs-cliniciens proposent d'autres méthodes (p. ex., un échantillon unique suivi dans le temps avec plusieurs mesures préintervention) afin de préconiser une approche, plus près de la réalité clinique. Il est aussi possible de constater que certains chercheurs utilisent de plus petits échantillons, mais les suivent de manière plus intensive, comme dans les études d'*experience sampling* (Granholm *et al.*, 2013) où les mêmes personnes répondent à des questionnaires brefs plusieurs fois par jour pendant une période déterminée assez brève (en général une semaine). Malgré l'ensemble de ces défis, l'ECR demeure une étape importante pour démontrer aux yeux de la communauté scientifique et clinique l'efficacité d'une intervention et ainsi favoriser son acceptation et son implantation. Pour finir, l'approche est rigoureuse, standardisée et comprend plusieurs étapes et règles préétablies permettant aux chercheurs intéressés de conduire un ECR selon les règles de l'art.

RÉFÉRENCES

CUTRONA, C.E. et D.W. RUSSEL (1987). « The provisions of social support and adaptation to stress », *Advance in Personal Relationships*, vol. 1, p. 37-67.

EDWARDS, J.R. et A.J. BAGLIONI (1993). « The measurement of coping with stress : Construct validity of the Ways of Coping Checklist and the Cybernetic Coping Scale », *Work and Stress*, vol. 7, p. 17-31.

FIRST, M.B., R.L. SPITZER, M. GIBBON et J.B.W. WILLIAMS (1997). *Structured Clinical Interview for DSM IVAxis I and II disorders – Patient Edition*, New York, Biometrics Research Department.

GOULD, R.A., K.T. MUESER, E. BOLTON, V. MAYS et D. GOFF (2001). « Cognitive therapy for psychosis in schizophrenia : An effect size analysis », *Schizophrenia Research*, vol. 48, p. 335-342.

GRANHOLM, E., D. BEN-ZEEV, D. FULFORD et J. SWENDSEN (2013). « Ecological Momentary Assessment of social functioning in schizophrenia : Impact of performance appraisals and affect on social interactions », *Schizophrenia Research*, vol. 13, <doi : 10.1016/j.schres.2013.01.005>.

JACOBSON, N.S. et P. TRUAX (1991). « Clinical significance : A statistical approach to defining meaningful change in psychotherapy research », *Journal of Consulting and Clinical Psychology*, vol. 59, p. 12-19.

LECOMTE, T., M. CORBIERE et F. LAISNÉ (2006). « Investigating self-esteem in individuals with schizophrenia : Relevance of the SERS », *Psychiatry Research*, vol. 143, p. 99-108.

LECOMTE, T., M. CYR, A.D. LESAGE, J.B. WILDE, C. LECLERC et N. RICARD (1999). « Efficacy of a self-esteem module in the empowerment of individuals with chronic schizophrenia », *Journal of Nervous and Mental Diseases*, vol. 187, p. 406-413.

LECOMTE, T., C. LECLERC, M. CORBIÈRE, T. WYKES, C.J. WALLACE et A. SPIDEL (2008). « Group cognitive behaviour therapy or social skills training for individuals with a first episode of psychosis ? Results of a randomized controlled trial », *Journal of Nervous and Mental Disease*, vol. 196, p. 866-875.

LECOMTE, T., C. LECLERC et T. WYKES (2012). « Group CBT for early psychosis – Are there still benefits one year later ? », *International Journal of Group Psychotherapy*, vol. 62, nº 2, p. 309-321.

LECOMTE, T., C. LECLERC, T. WYKES et J. LECOMTE (2003). « Group CBT for clients with a first episode of psychosis », *Journal of Cognitive Psychotherapy : An International Quarterly*, vol. 17, p. 375-384.

LUKOFF, D., K. NUECHTERLEIN et J. VENTURA (1986). « Manual for the Expanded Brief Psychiatric Rating Scale », *Schizophrenia Bull*etin, vol. 12, p. 578-602.

MELDRUM, M.L. (2000). « A brief history of the randomized controlled trial. From oranges and lemons to the gold standard », *Hematology Oncologu Clinical North America*, vol. 14, nº 4, p. 745-760.

MUESER, K.T., D.L. NOORDSY, R.E. DRAKE et L. FOX (2001). « Integrated treatment for severe mental illness and substance abuse : Effective components of programs for persons with co-occurring disorders », *Santé mentale au Quebec*, vol. 26, n° 2, p. 22-46, <doi :014524ar> [pii].

SCHULZ, K.F., D.G. ALTMAN et D. MOHER (2010). « CONSORT 2010 Statement : Updated guidelines for reporting parallel group randomised trials », *Annals of International Medicine*, vol. 152, Epub, 24 mars.

SENSKY, T., D. TURKINGTON, D. KINGDON, J.L. SCOTT, R. SIDDLE, M. O'CARROLL *et al*. (2000). « A randomized controlled trial of cognitive-behavioral therapy for persistent symptoms in schizophrenia resistant to medication », *Archives of General Psychiatry*, vol. 57, p. 165-172.

SPIDEL, A., T. LECOMTE et C. LECLERC (2006). « Community implementation successes and challenges of a cognitive-behavior therapy group for individuals with a first episode of psychosis », *Journal of Contemporary Psychotherapy*, vol. 36, p. 51-58.

TARRIER, N. et T. WYKES (2004). « Is there evidence that cognitive behaviour therapy is an effective treatment for schizophrenia ? A cautious or cautionary tale ? », *Behaviour Research and Therapy*, vol. 42, p. 1377-1401.

VENTURA, J., M.F. GREEN, A. SHANER et R.P. LIBERMAN (1993). « Training and quality assurance with the Brief Psychiatric Rating Scale », *International Journal of Methods in Psychiatric Research*, vol. 3, p. 221-244.

WALLACE, C.J. (2001). *Skills Training Fidelity Scale*, Los Angeles, UCLA Psychiatric Rehabilitation Center.

YOUNG, J. et A. BECK (1980). *Cognitive Therapy Scale Rating Manual*, Philadelphie, Center for Cognitive Therapy.

CHAPITRE 11

LES ÉVALUATIONS ÉCONOMIQUES DANS LE DOMAINE DE LA SANTÉ
Illustration d'une analyse des coûts du programme Integrated Psychological Treament (IPT)

Pierre-Alexandre Dionne
Helen-Maria Vasiliadis

FORCES

- Elles permettent l'obtention de données probantes pertinentes pour la prise décisionnelle et pour le développement de politiques de santé.
- L'analyse coût-utilité apporte notamment un indicateur d'efficience standardisé facilement comparable entre diverses interventions en santé, facilitant ainsi l'allocation des ressources.
- Elles utilisent des modèles statistiques avancés permettant de modéliser la progression d'une cohorte d'individus sans nécessairement avoir à suivre ces derniers sur de très longues périodes.

LIMITES

- Plusieurs limites méthodologiques sont reliées à l'incertitude des données de coûts et d'effets.
- Les méthodes d'évaluations économiques sont complexes et non intuitives pour plusieurs experts et décideurs du domaine de la santé.
- Les méthodes de monétarisation de la vie humaine entraînent certains enjeux éthiques.

L'évaluation économique, partie intégrante du processus décisionnel en santé, a pris un essor important au cours des dernières décennies. Ce type d'évaluation renseigne sur les coûts de la maladie et des interventions cliniques et psychosociales, ainsi que sur leurs bénéfices cliniques et économiques. L'évaluation économique permet ainsi de fournir des données probantes visant à favoriser le développement de politiques de santé, un meilleur mode d'offre des services de santé et une allocation adéquate des ressources. Ce chapitre présentera le champ de l'économie de la santé et les méthodes utilisées. De plus, il proposera un exemple concret d'étude économique touchant le domaine de la santé mentale.

1. DÉFINITION, VISÉES ET PARTICULARITÉS DE L'ÉCONOMIE DE LA SANTÉ

Avec le vieillissement de la population et la hausse importante du coût des nouvelles avancées technologiques et des traitements, les budgets en santé ne cessent de croître, principalement dans les pays occidentaux. Pour les décideurs du système de santé, cette nouvelle réalité entraîne un important questionnement concernant l'efficacité (c.-à-d. l'atteinte des objectifs visés, et ce, tout en démontrant sécurité, réactivité, accessibilité) et surtout l'efficience (c.-à-d. l'atteinte d'un niveau optimal d'effets pour les ressources investies dans les soins de santé) (Brousselle *et al.*, 2009 ; Drummond *et al.*, 2005). Par exemple, est-il acceptable de prescrire un nouveau médicament pour la dépression si son coût est nettement supérieur avec des effets cependant légèrement supérieurs au médicament prescrit actuellement ? Vaut-il la peine d'investir en prévention des maladies mentales pendant que les ressources nous empêchent d'ouvrir un certain nombre de lits d'hôpital et de former davantage de médecins de famille ? L'économie de la santé se penche sur ces questions et tente d'y répondre.

L'économie de la santé est définie comme l'application de la science économique au domaine de la santé qui a pour objectif d'évaluer la valeur, l'efficience, l'efficacité et le comportement associés à la consommation et à la production de la santé et des soins de santé (Brousselle *et al.*, 2009 ; Prince *et al.*, 2003). Toutefois, cette discipline présente plusieurs distinctions avec les autres domaines de l'économie, soit l'intervention des gouvernements dans le domaine de la santé, les barrières de l'accès aux soins et la présence d'un tiers payeur ou décideur (c.-à-d. il est de la responsabilité du professionnel de la santé de choisir la thérapie appropriée pour le patient, et les soins sont généralement couverts par des régimes d'assurance publics ou privés) (Evans, 1997 ; Phelps, 2003). À l'inverse des autres sciences économiques,

l'économie de la santé ne suit pas les lois du libre marché; elle est plutôt guidée par des décisions politiques et institutionnelles pour l'allocation des ressources de santé (Brousselle *et al.*, 2009; Evans, 1997).

1.1. Terminologie et classification des études économiques

1.1.1. Terminologie des coûts et effets en santé

Coûts et effets directs médicaux

Cette terminologie est employée pour décrire les coûts et les effets associés à la santé et à l'utilisation des services de santé. Cette catégorie comprend l'ensemble des ressources matérielles (médications, instruments médicaux) et humaines (salaires des professionnels du système de la santé) reliées directement au diagnostic, au traitement et au suivi du patient (Drummond *et al.*, 2005).

Coûts et effets directs non médicaux

Les coûts directs non médicaux comprennent l'ensemble des coûts engendrés par l'utilisation des services de santé sans toutefois être directement reliés au traitement ou au suivi médical du patient. Il s'agit majoritairement de frais généraux, dont font partie la buanderie, l'entretien et l'administration d'établissements de santé. Dans cette catégorie, certains auteurs considèrent les frais assumés par le patient pour le transport reliés à la visite médicale (Drummond *et al.*, 2005).

Coûts et effets indirects

Il s'agit en quelque sorte des conséquences indirectes d'une intervention ou d'une maladie, notamment les coûts associés à la mortalité prématurée et à la perte de productivité. L'absentéisme au travail résultant d'une condition médicale entraînera une perte de productivité sur le marché du travail. Cette perte de productivité peut être évaluée et transformée en valeur marchande selon les méthodes du capital humain ou du coût de friction[1] (Drummond *et al.*, 2005).

1. La méthode du capital humain permet d'accorder une valeur marchande à la perte de productivité en quantifiant le nombre d'heures d'absence au travail, tandis que la méthode du coût de friction évalue uniquement le nombre d'heures d'absence au travail avant qu'un autre employé remplace l'individu concerné. Le livre de Drummond et collaborateurs (2005) permet une lecture plus approfondie sur ces méthodes.

Les coûts et les effets à répertorier seront déterminés en fonction de la perspective de l'étude. Il peut s'agir du point de vue d'un centre hospitalier, du système de santé ou de la société dans son ensemble. Par exemple, si une étude évalue une technologie médicale selon la perspective d'un centre hospitalier, les coûts associés à l'absentéisme au travail ne seront pas répertoriés.

1.1.2. Classification des études économiques en santé

Les études économiques sont classifiées en deux principales catégories d'études : les évaluations économiques partielles et les évaluations économiques complètes. Pour être considérée comme complète, une évaluation économique doit comparer les coûts et les effets, et doit effectuer la comparaison entre deux ou plusieurs comparateurs (approche différentielle).

Les *évaluations économiques partielles* sont des études à visée descriptive qui évaluent les coûts associés à une intervention, à une problématique, à une maladie. Parmi les études les plus utilisées, les études du coût de la maladie (*cost-of-illness*) visent à décrire l'ampleur des conséquences d'un problème de santé pour prioriser l'établissement de mesures concrètes et soulever des pistes de recherche futures (Tarricone, 2006).

Les *évaluations économiques complètes* estiment la valeur économique d'un traitement ou d'une intervention en comparant ses coûts et ses effets avec ceux d'un comparateur (p. ex., analyse coût-efficacité). Ainsi, le choix du comparateur est très important dans l'optique d'évaluer correctement le coût-efficacité d'un traitement et d'apporter des données pertinentes pour la prise de décision (Drummond *et al*., 2005). En général, le comparateur est le traitement ou l'intervention couramment utilisé pour la même indication et préférablement l'option de référence (*standard of care*) dans la classe thérapeutique. Ces études sont devenues des outils primordiaux à la prise de décision dans plusieurs pays, notamment dans le contexte de systèmes de santé financés par l'État.

1.2. Méthodes d'évaluations économiques

Il existe cinq types d'évaluations économiques en santé : 1) les analyses coût-minimisation, 2) coût-efficacité, 3) coût-utilité, 4) coût-bénéfice, et 5) coût-conséquence. Comme présentées au tableau 11.1, ces études se différencient principalement sur le plan de l'évaluation des effets sur la santé et de leur monétarisation (aussi appelée valorisation). Le choix de l'analyse dépend de la nature de l'intervention et du comparateur choisis (Brouselle *et al*., 2009 ; Drummond *et al*., 2005).

1.2.1. Analyse coût-minimisation (ACM)

L'ACM compare deux interventions ayant exactement les mêmes effets, mais des coûts différents. Ainsi, l'option la moins chère sera à privilégier.

1.2.2. Analyse coût-efficacité (ACE)

L'ACE est utilisée pour comparer deux interventions qui ont un objectif unique et non ambigu, mais une efficacité et des coûts différents. Ce type d'analyse mesure les effets d'un programme ou d'une intervention en unités physiques (p. ex., le nombre de cas dépistés, les années de vie sauvées, le nombre de jours d'absentéisme évités, la diminution dans les symptômes de détresse psychologique avec l'échelle de Kessler [K-10]) (Kessler *et al.*, 2003). Les résultats sont présentés sous forme d'un ratio coût-efficacité incrémental (RCEI), par exemple 5000 $ par cas traité.

RCEI ($/unité physique d'efficacité) = (C2 – C1)/(E2 – E1)

où

C = coûts du programme ou de l'intervention et tous les autres coûts associés;

E = effets du programme ou de l'intervention en unités physiques.

1.2.3. Analyse coût-utilité (ACU)

L'ACU permet de comparer des interventions avec plusieurs effets de nature et de gravité différentes, et de tenir compte de la qualité de vie de l'individu en utilisant une seule et unique mesure: l'utilité. L'utilité, concept central du calcul d'années de vie ajustées pour la qualité (Quality-Adjusted Life Years – QALY), est une mesure de préférence ou de satisfaction envers un état de santé précis en fonction de la qualité de vie résultante sur une échelle numérique de 0 à 1, où 0 est la mort et 1 est une parfaite santé (Drummond *et al.*, 2005). L'utilité est mesurée à l'aide d'outils spécialisés: les échelles, l'arbitrage temporel, le pari standard et les systèmes de classification multiattributs (Drummond *et al.*, 2005).

Les *échelles visuelles* se présentent sous la forme d'une ligne verticale graduée sur laquelle l'individu doit classer différents états de santé selon leur degré de gravité et la satisfaction qu'il ressentirait en passant d'un état à l'autre (Crochard-Lacour et LeLorier, 2000). Un exemple serait l'Échelle visuelle analogique (Visual Analogue Scale), où l'individu évalue son état de santé sur une échelle de 0 (la mort) à 100 (la parfaite santé).

Le *pari standard* consiste à demander à l'individu de choisir entre un état de santé donné certain ou une situation ayant deux issues possibles, soit la mort et la bonne santé (Crochard-Lacour et LeLorier, 2000). Dans cet exercice, on demande au sujet s'il préfère continuer à vivre dans son état actuel d'incapacité ou s'il est prêt à accepter un traitement qui peut lui redonner une parfaite santé, mais qui a aussi un risque de causer la mort. Dans ce cas, le risque de décès peut varier et l'individu indique le risque de décès qu'il est prêt à accepter afin d'être en parfaite santé.

L'*arbitrage temporel* consiste à demander à l'individu de déterminer la durée de vie en bonne santé qu'il juge équivalente à son espérance de vie dans un état de santé donné (Crochard-Lacour et LeLorier, 2000).

Finalement, à propos des *systèmes de classification multiattributs*, les principaux outils de mesure d'utilité de nos jours permettent la conversion directe des résultats d'un questionnaire de qualité de vie en mesure d'utilité (p. ex., Short-Form Health Survey–SF-6D [Brazier, Roberts et Deverill, 2002], EuroQol – EQ-5D [The EuroQol Group, 1990], Heath Utilities Index – HUI-3 [Feeny *et al.*, 1995]). Ces outils sont administrés sous forme de questionnaires autoadministrés évaluant l'état de santé du sujet sur différentes dimensions (p. ex., pour l'EQ-5D, il y a cinq dimensions: mobilité, soins personnels, activités courantes, douleur-inconfort, anxiété-dépression) comprenant chacune plusieurs niveaux de gravité (The EuroQol Group, 1990). Par la suite, un poids relatif à chaque attribut (résultat sur le questionnaire) est appliqué dans la population générale et le score final obtenu est l'utilité. Une fois la valeur d'utilité évaluée, il est possible de calculer le nombre de QALY (Années de vie ajustées pour la qualité * Utilité) pour ainsi obtenir une unité de mesure standardisée. Le ratio coût-utilité incrémental (RCUI) associé est une mesure d'efficience qui permet la comparaison directe de plusieurs interventions facilite ainsi le processus décisionnel pour l'adoption de nouvelles technologies ou de nouveaux traitements médicaux.

$$\text{RCUI (\$/QALY)} = (C_2 - C_1)/(U_2 * \text{Années de vie} - U_1 * \text{Années de vie})$$

où

C = coûts du programme ou de l'intervention;

U = mesure d'utilité associée à l'intervention.

1.2.4. Analyse coût-bénéfice

Aussi appelé coût-avantage, ce type d'analyse vise à évaluer la rentabilité des interventions en comparant les coûts et les effets, où l'ensemble des effets (bénéfices ou conséquences) est converti en valeur marchande afin

de permettre une comparaison directe avec les coûts de l'intervention. Par exemple, pour un programme de prévention du suicide qui a des effets positifs sur les taux de suicide dans une population, l'évaluation coût-bénéfice va comparer les coûts associés au programme de prévention à ceux des coûts évités en fait de mortalité évitée. Ainsi, une valeur marchande doit être attribuée à chaque année de vie excédentaire due au programme. Pour ce faire, deux principales approches permettent d'accorder une valeur marchande à la vie humaine: l'approche du capital humain et l'évaluation contingente (Drummond *et al.*, 2005). L'approche du capital humain consiste à valoriser le temps de vie d'un individu par sa valeur productive, soit généralement le montant de son salaire ou le revenu national moyen par habitant (Crochard-Lacour et LeLorier, 2000). L'évaluation contingente est basée sur la propension à payer d'un individu pour qu'un traitement lui soit administré ou à accepter une compensation pour renoncer à un traitement (Crochard-Lacour et LeLorier, 2000). Toutefois, ces méthodes apportent leurs lots de critiques, du point de vue autant méthodologique qu'éthique (Brouselle *et al.*, 2009; Drummond *et al.*, 2005). L'analyse coût-bénéfice est principalement utilisée dans le domaine de la santé publique et de l'évaluation de programmes.

Ratio coût-bénéfice = C/E

Bénéfice net = E-C

où

C = coûts totaux du programme ou de l'intervention;

E = tous les effets et externalités du programme ou de l'intervention convertis en valeur marchande.

1.2.5. Analyse coût-conséquence

Similaire à l'analyse coût-bénéfice, elle mesure l'ensemble des coûts et des conséquences associés à une intervention, dont les effets cliniques, les effets indésirables et les externalités (effets indirects). Toutefois, elle énumère l'ensemble des effets sous forme de leur unité naturelle sans les convertir sous une même unité, permettant ainsi aux décideurs d'analyser eux-mêmes les données. Bien qu'elle tienne compte de l'hétérogénéité des effets d'une intervention et évite l'emploi de techniques complexes, cette analyse est rarement utilisée, car elle ne permet pas d'obtenir un indicateur d'efficience standardisé, ce qui rend plus complexe le processus décisionnel (CADTH, 2006; Drummond *et al.*, 2005).

Tableau 11.1.
Sommaire des types d'évaluations économiques en économie de la santé

Types d'évaluations économiques	Coûts des interventions	Caractérisation des effets	Mesure et évaluation des effets	Unités et critères décisionnels
Coût-minimisation	$	Identiques entre les options	Aucune mesure nécessaire	Coûts minimaux = meilleure option
Coût-efficacité (ACE)	$	Un seul effet mesuré	Unité naturelle (p. ex., cas dépisté, symptômes de détresse psychologique)	$/unité naturelle
Coût-utilité (ACU)	$	Effets multiples sur la santé et la qualité de vie	Transformation en Années de vie ajustées pour la qualité (QALY)	$/QALY
Coût-bénéfice (ACB)	$	Effets multiples sur la santé et la qualité de vie	Monétarisation de tous les effets ($)	Bénéfices nets (B-C) Ratio B/C
Coût-conséquence	$	Effets multiples	Effets décrits selon leurs unités naturelles	Jugement du décideur

Source : Adapté de Drummond *et al.*, 2005.

1.2.6. Fondements méthodologiques d'une évaluation économique

Avant de mener une évaluation économique, il est important de bien connaître l'intervention dans son ensemble : les effets directs, les externalités, la structure de financement et tout autre aspect connexe, tel que l'épidémiologie de la maladie concernée. De plus, l'auteur doit choisir un *comparateur*, une *population cible* et une *perspective appropriée* (CADTH, 2006). La perspective de l'étude influencera le choix des coûts et les effets à considérer. De plus, l'auteur doit utiliser un *horizon temporel* adéquat pour répertorier l'ensemble des répercussions futures du traitement ou de la condition médicale concernée (CADTH, 2006). Pour ce faire, plusieurs techniques de modélisation sont utilisées, dont le modèle de Markov[2] (Drummond *et al.*, 2005 ; Sonnenberg et Beck, 1993).

2. Le modèle de Markov est un modèle décisionnel complexe qui permet de représenter plus fidèlement la progression d'une cohorte qu'un arbre décisionnel simple. Pour une lecture approfondie concernant l'analyse décisionnelle et le modèle de Markov, le livre de Drummond et collaborateurs (2005) présente ces notions et leurs applications tandis que l'article de Sonnenberg et Beck (1993) présente des notions plus avancées et théoriques.

La projection des effets selon un horizon temporel supérieur à un an implique l'*actualisation* des coûts et des effets. Concept distinct de celui de l'inflation, l'actualisation permet de refléter les préférences temporelles des individus (Drummond *et al.*, 2005). Une somme d'argent a une plus grande valeur aujourd'hui que la même somme d'argent dans le futur. Il n'y a pas de consensus précis concernant le taux d'actualisation ; il varie généralement entre 3 % et 10 % selon les recommandations de certains pays ou de certaines agences spécialisées en évaluation économique (Smith et Gravelle, 2000). Selon l'Agence canadienne d'évaluation des technologies en santé, le taux de référence d'actualisation est de 5 % au Canada (CADTH, 2006).

Les évaluations économiques présentent généralement de nombreuses limites, notamment sur le plan de la mesure des coûts (p. ex., utilisation de données provenant de différents pays et contextes cliniques, données manquantes, données présentant un important niveau d'incertitude) et sur celui des effets (p. ex., certains effets adverses non répertoriés et quantifiés, données d'efficacité provenant d'études ayant des limites méthodologiques, technique d'évaluation de l'utilité inadéquate, données provenant d'opinions de différents experts). Il est donc important de bien délimiter les sources d'incertitude dans les données utilisées pour l'évaluation économique et d'effectuer des *analyses de sensibilité* pour évaluer la robustesse des résultats. Différentes analyses de sensibilité existent et elles sont regroupées en deux principales catégories : analyses de sensibilité déterministes univariées (c.-à-d. l'auteur fait varier manuellement les paramètres qui présentent un certain degré d'incertitude) et analyses de sensibilité probabilistes (c.-à-d. l'auteur fait varier plusieurs paramètres simultanément à l'aide de techniques de simulation tels que la méthode de Monte-Carlo par chaîne de Markov[3]) (Drummond *et al.*, 2005).

En résumé, les évaluations économiques sont des analyses très utiles pour le processus décisionnel en santé, mais elles sont toutefois des analyses complexes et susceptibles de présenter plusieurs limites. Bien que certaines techniques permettent d'évaluer la robustesse des résultats et leur sensibilité à des variations sur le plan de certains paramètres, il est de la responsabilité du chercheur de démontrer rigueur et transparence vis-à-vis des limites de son étude. Différentes lignes directrices pour la tenue d'évaluations économiques ont été développées par des agences gouvernementales ou par des experts du domaine (CADTH, 2006 ; Drummond *et al.*, 2005). Une

3. La méthode de Monte-Carlo par chaîne de Markov est une méthode d'analyse probabiliste combinée au modèle de Markov qui permet d'étudier l'interaction entre les distributions de probabilités rattachées à chaque variable afin de fournir une estimation du degré d'incertitude d'un modèle (définition tirée du glossaire *HTaglossary.net*). Le livre de Drummond et collaborateurs (2005) permet une lecture plus approfondie sur l'analyse de sensibilité probabiliste et la méthode de Monte-Carlo.

adaptation de la grille d'évaluation de la qualité d'une évaluation économique de Drummond et collaborateurs est présentée à l'annexe 11.1. Les principaux pièges rencontrés lors de la tenue d'une évaluation économique et lors de la publication des résultats sont la spécification inadéquate des paramètres utilisés pour la modélisation (p. ex., perspective de l'étude ou horizon temporel inadéquat, utilisation de mauvais comparateurs, aucune actualisation des coûts et des effets), la gestion inadéquate de l'incertitude dans les sources de données (p. ex., absence d'analyses de sensibilité ou analyses de sensibilité inadéquates), et le manque de transparence dans la publication des résultats (p. ex., données non présentées, la discussion n'aborde pas les limites de l'étude) (Drummond *et al.*, 2005).

1.2.7. Critères décisionnels et notions d'acceptabilité

Les résultats des évaluations économiques ne permettent pas à eux seuls de conclure sur le coût-efficacité d'un traitement ou d'une intervention ; tout est question d'acceptabilité et de disposition à payer. Les ratios RCEI ou RCUI ci-dessus présentés doivent être analysés en considérant qu'il s'agit d'indices d'efficience incrémentale entre deux traitements et non d'indices d'efficience absolue (Drummond *et al.*, 2005). Étant donné qu'une nouvelle technologie plus efficace est rarement moins coûteuse que l'ancienne technologie, le décideur doit se demander s'il est acceptable de payer un prix excédentaire pour les bénéfices obtenus.

Bien qu'il n'existe aucun seuil précis pour statuer de l'acceptabilité, plusieurs auteurs ont toutefois pu établir à partir de quel seuil de RCUI une nouvelle intervention est susceptible d'être adoptée (Grosse, 2008). Il a été soulevé que plusieurs agences gouvernementales ont tendance à accepter le remboursement de traitements avec un RCUI inférieur à 50 000 $ par QALY (Grosse, 2008). Toutefois, ces seuils d'acceptabilité sont des figures arbitraires qui varient en fonction de plusieurs facteurs, dont le contexte économique d'un pays et la disponibilité des ressources, la perspective de l'étude, la population à l'étude, etc. (Grosse, 2008 ; Owens, 1998). Enfin, bien que le critère d'efficience soit très important dans le domaine de la santé, il ne faut pas oublier que la prise de décision doit ultimement respecter d'autres valeurs très importantes dont l'équité et le respect des libertés individuelles (Contandriopoulos, 2005).

1.3. Logiciels d'analyse

Pour l'analyse des coûts, les logiciels statistiques (p. ex., SAS, SPSS, STATA) sont tout à fait adéquats. Lors du choix de l'analyse statistique, il est important de considérer que les coûts de santé ont généralement la particularité de ne pas suivre une distribution normale et de ne pas toujours satisfaire les critères d'homoscédasticité (c.-à-d. le même degré de variation des résidus pour chaque observation de la variable dépendante X). Dans le cas de données qui ne suivent pas une distribution normale, les tests paramétriques traditionnels (p. ex., chi-carré, test *t*) ne peuvent pas être utilisés. L'expérimentateur doit choisir entre des tests statistiques non paramétriques, tels que le test des rangs signés de Wilcoxon (Armitage, Berry et Matthews, 2002), ou effectuer une transformation pour tenter de normaliser les données. Une procédure statistique avancée a été proposée par Manning et Mullahy (2001) pour choisir la transformation adéquate des données avant d'effectuer un test paramétrique de type modèle linéaire généralisé. Cette procédure consiste à choisir le bon test statistique et la bonne spécification du modèle linéaire généralisé (c.-à-d. considérant une distribution normale, de poisson ou gamma) par le moyen d'une transformation Box-Cox et d'un test de Park modifié (Manning et Mullahy, 2001). De plus, les évaluations économiques peuvent être effectuées à l'aide du logiciel Microsoft Excel. D'autres logiciels ont été spécialement développés pour la modélisation plus avancée et l'analyse décisionnelle comme TreeAge Pro (TreeAge Software Inc.) et Crystal Ball (Oracle Corporation, Redwood Shores, CA). Ces logiciels sont des compléments du logiciel Excel qui permettent à la fois de faciliter la modélisation et les analyses de sensibilité probabilistes et d'illustrer l'arbre décisionnel et les résultats de manière plus esthétique.

2. ILLUSTRATION D'UNE ANALYSE DESCRIPTIVE DES COÛTS D'UN PROGRAMME D'INTERVENTION EN SANTÉ MENTALE

Un exemple pertinent d'analyse de l'utilisation des services de santé et des coûts associés à une intervention concerne le programme Integrated Psychological Treatment (IPT). À la suite de l'introduction d'une version québécoise de l'IPT dans des établissements spécialisés pour le suivi de personnes atteintes de schizophrénie, des auteurs ont évalué non seulement les effets cliniques, mais également l'utilisation des services de santé et les coûts associés. L'analyse descriptive des coûts (évaluation économique partielle) reliés à l'utilisation des soins de santé associés au programme IPT, parue dans le *Journal of Mental Health Policy and Economics* (Vasiliadis *et al.*, 2006), est présentée dans cette section.

2.1. Utilisation des services de santé associés au programme IPT pour personnes atteintes de schizophrénie

La schizophrénie est un trouble psychiatrique très invalidant ; elle est associée à plusieurs comorbidités, a des répercussions importantes sur la productivité et engendre des coûts de soins de santé considérables (WHO, 2003). Au Canada, les coûts de la schizophrénie ont été estimés à 6,85 milliards de dollars en 2004 (Goeree *et al.*, 2005). Étant donné ces coûts importants et les nombreuses restrictions actuelles des dépenses de santé, il est important de bien évaluer les coûts des interventions et des programmes qui visent la prise en charge et le rétablissement des personnes aux prises avec une schizophrénie. C'est notamment le cas pour le programme IPT (Brenner *et al.*, 1994) ; bien que ce programme ait été démontré efficace pour améliorer la qualité de vie, le fonctionnement cognitif et social ainsi que l'accomplissement des activités quotidiennes dans plusieurs études au Québec et ailleurs dans le monde (Brenner *et al.*, 1992 ; Mueller, Roder et Brenner, 2003 ; Briand *et al.*, 2006), les coûts qui y étaient associés n'avaient pas encore été décrits.

L'objectif de l'étude de Vasiliadis *et al.* (2006) était de décrire et de comparer l'utilisation des ressources de soins de santé et les coûts supportés par les personnes ayant une schizophrénie au cours de leur participation au programme IPT, ainsi que ceux supportés dans l'année qui a précédé leur intégration dans différents milieux cliniques.

2.2. Méthodologie

2.2.1. Population et horizon temporel de l'étude

Des personnes aux prises avec une schizophrénie, selon les critères du DSM-IV (APA, 2000), jugées stables et capables de poursuivre une intervention de groupe ont été recrutées dans neuf établissements de santé servant de centres d'étude. Au total, 87 des 90 patients recrutés ont participé à au moins deux séances du programme IPT. L'âge moyen de la cohorte était de 32,7 ans (é.t. 11,9) et 72 % étaient des hommes (Briand *et al.*, 2006). Sur les 87 patients, 55 ont terminé le programme ; ceux qui ont cessé leur participation (*n* = 32) ont consenti à ce que leurs dossiers médicaux soient étudiés pour la durée totale de l'étude. L'étude a inclus un suivi rétrospectif (de 10 à12 mois avant le début du programme IPT) et un suivi durant toute la durée du programme (de 10 à 19 mois).

2.2.2. Programme intégratif de thérapies psychologiques (IPT)

L'IPT est un programme cognitif comportemental, conçu pour des groupes de 8 à 10 personnes aux prises avec une schizophrénie se réunissant deux fois par semaine pendant environ une année. Le programme, administré par des professionnels en santé mentale (majoritairement des ergothérapeutes), se compose d'environ 85 séances d'une heure et demie. Le programme comprend six sous-programmes hiérarchisés, soit la différenciation cognitive, la perception sociale, la communication verbale, les aptitudes sociales, la gestion des émotions et la résolution de problèmes, comprenant chacun trois domaines de complexité croissante (Brenner *et al.*, 1994). Des séances dans les milieux naturels (hors des cliniques) ont été ajoutées à la version québécoise de l'IPT pour aider les personnes aux prises avec une schizophrénie à transférer dans leur contexte de vie réelle les habiletés et les stratégies acquises.

2.2.3. Analyse des coûts

L'analyse descriptive des coûts a été effectuée selon la perspective du système de santé et des services sociaux québécois (coûts médicaux et coûts indirects) et la perspective du patient (coûts non médicaux directs). Les coûts supportés ont été calculés et présentés pour deux périodes de temps ; les coûts générés dans l'année précédant le programme ainsi que ceux générés au cours du programme. La mesure et l'évaluation des coûts des services de santé et des ressources ont été basées sur les lignes directrices en économie de la santé (CCOHTA, 1996 ; CADTH, 2006 ; IHE, 2000).

Les coûts médicaux directs ont été comptabilisés en fonction des ressources utilisées pendant une hospitalisation, une visite à l'urgence (ED), une visite ambulatoire, des honoraires des médecins et des médicaments utilisés. Les coûts liés au logement ont également été estimés. Les coûts non médicaux directs et indirects de la perspective du patient représentaient le transport et la perte de productivité (temps perdu) pour la recherche de soins médicaux ambulatoires.

2.2.4. Mesure de l'utilisation des ressources de santé

L'utilisation des ressources de soins de santé a été mesurée à partir de diverses sources. L'information sur les hospitalisations (durée de séjour, nombre et type de soins) recueillie à partir des dossiers médicaux a été complétée et validée par les fichiers administratifs de Maintenance et exploitation des données pour l'étude de la clientèle hospitalière (MED-ÉCHO). Ces fichiers contiennent les données médico-hospitalières et cliniques qui portent sur

les patients hospitalisés dans les établissements québécois de soins généraux et spécialisés, et comprennent le diagnostic, le nombre de visites, la durée d'hospitalisation et les services médicaux rendus. Les interventions médicales pratiquées dans les hôpitaux étaient fournies par la Régie de l'assurance maladie du Québec (RAMQ). Le nombre de consultations ambulatoires en médecine générale et en psychiatrie ont également été recueillies à partir des dossiers médicaux des patients et des fichiers de la RAMQ. Ces deux sources d'information ont été appariées afin d'éviter la présence d'un biais d'information pouvant sous-estimer l'utilisation réelle des services médicaux et psychosociaux en cliniques externes. L'utilisation des médicaments a été estimée à partir du registre des services pharmaceutiques de la RAMQ, qui détient les informations relatives aux médicaments distribués pour chaque patient.

2.2.5. Évaluation des coûts des ressources de soins de santé

Pour estimer le coût direct moyen des ressources de santé utilisées, les rapports financiers des établissements de santé ont été consultés pour calculer le coût unitaire des services et ressources provenant d'un certain centre d'activité (CA) dans les établissements concernés (MSSS, 2001). Les coûts unitaires moyens pour un CA comprenaient également les coûts de soutien (c.-à-d. sécurité, administration, buanderie et autres frais généraux). Comme les patients ont été recrutés dans différents centres hospitaliers, les coûts moyens ont été obtenus à partir de chaque hôpital, ce qui a ajouté de la précision à l'estimation. Tous les coûts sont ceux de l'exercice financier 2000-2001. Le tableau 11.2 résume les mesures, les sources de données et les méthodes d'évaluation des coûts pour chacune des catégories de services de santé considérées dans cette étude.

2.2.6. Services ambulatoires

Les services ambulatoires de santé mentale ont généralement été effectués en cabinets privés ainsi que dans les cliniques externes. Les services psychiatriques ambulatoires incluent ceux octroyés par des médecins (psychiatres) (données des fichiers médicaux de la RAMQ) ainsi que d'autres professionnels paramédicaux (ergothérapeutes) et psychosociaux (psychologues, travailleurs sociaux) (données incluses dans le coût unitaire moyen calculé à partir des rapports financiers des établissements). Chaque visite a été évaluée sur la base du coût moyen par heure dans un certain département. Dans notre étude, les visites individuelles à la clinique externe et les réunions de groupe ont duré en moyenne 45 et 120 minutes respectivement. Une visite chez l'omnipraticien ou le médecin spécialiste a également été confirmée

Tableau 11.2.
Principales catégories de coûts considérées pour l'analyse des coûts et leurs principales variables, mesures et sources de données

Catégories de soins de santé	Services de santé répertoriés	Mesures	Sources de données	Évaluation en $
A : Perspective du système de santé				
1. Visites médicales ambulatoires	Visites en salles d'urgence	▪ Nombre de visites ▪ Durée de séjour ▪ Coûts	▪ Fichiers administratifs RAMQ (consultations médicales) ▪ Fichiers MED-ÉCHO (visites à l'urgence) ▪ Dossiers médicaux des participants	▪ Rapports financiers des établissements de santé (visites à l'urgence : $/visite ; professionnels en santé mentale : $/heure) (n'incluent pas les honoraires des médecins) ▪ Fichiers RAMQ : frais d'actes médicaux
	Visites ambulatoires autres (cabinet privé, centre local de services communautaires et clinique externe)	▪ Nombre de visites ▪ Type de soins (généraux ou en psychiatrie) ▪ Coûts		
2. Visites en milieux hospitaliers	Hospitalisations	▪ Nombre ▪ Durée de séjour ▪ Type de soins ▪ Coûts	▪ Fichiers administratifs MED-ÉCHO ▪ Dossiers médicaux des patients	▪ Rapports financiers des établissements ($ par jour ; $ par acte diagnostic (sauf honoraires des médecins) ▪ Fichiers RAMQ : frais d'actes médicaux
3. Médicaments	Médicaments couverts par l'assurance médicaments de la RAMQ	Coûts	Registre des services pharmaceutiques de la RAMQ	$ des médicaments de prescription (incluant les frais de service des pharmaciens)
4. Actes médicaux	Facturation des actes médicaux (honoraires des médecins)	Coûts	Données de facturation médicale de la RAMQ	$ totaux pour les actes médicaux
5. Frais de logement	Logements subventionnés par le MSSS	Coûts	Fichiers administratifs du MSSS	$ des paiements quotidiens du MSSS octroyés aux résidences
B : Perspective du patient				
1. Coûts de transport		Coûts	Données de l'Agence du revenu du Canada	$ de transport/km
2. Coûts de perte de productivité	Temps pris pour la recherche de soins médicaux	Nombre d'heures et coûts	Statistique Canada, taux horaire pour emploi rémunéré	$/heure reliés au temps passé à rechercher des soins médicaux
3. Frais de logement	Logements subventionnés par le MSSS	Coûts	Entrevue avec les patients	Contribution du patient aux frais de logement ($ mensuel)

à l'aide des données de facturation médicale de la RAMQ. L'évaluation du coût des ressources utilisées au cours de ces consultations ambulatoires a été réalisée de manière similaire.

Pour l'IPT, les cliniciens spécialisés en santé mentale ont été formés en moyenne trois journées de huit heures. Au terme de l'étude, il y avait quatre heures de conversation de suivi avec le coordonnateur de l'étude et deux séances finales de deux heures chacune. La formation des professionnels a été évaluée en fonction du coût moyen par heure travaillée dans leur établissement (centre d'activité). Pour l'IPT, l'évaluation des ressources associées à chaque séance a été effectuée de manière similaire à celle des visites psychiatriques en cliniques ambulatoires.

Une visite à l'urgence a été répertoriée une fois que le patient était inscrit à l'accueil. Dans les rapports financiers des CA des départements d'urgence, un coût moyen par visite était rapporté. Le coût moyen par visite incluait le salaire du personnel, les instruments médicaux et fournitures de bureau ainsi que les autres frais généraux (administration, entretien, etc.). Les coûts de la pharmacie, les coûts de nutrition et toute autre consultation avec d'autres professionnels ont également été comptabilisés (voir section Services de soins hospitaliers).

2.2.7. Services de soins hospitaliers

Une hospitalisation a été évaluée sur la base d'un coût par jour. Les données des CA ont fourni de l'information sur les ressources pour tous les soins de traitement et de suivi donnés aux patients pendant leur séjour, les salaires, les équipements médicaux et les fournitures de bureau et autres frais généraux. Les salaires visent uniquement ceux reliés aux soins infirmiers. Les services psychosociaux et paramédicaux ont été mesurés dans le contexte des soins ambulatoires (voir section Services ambulatoires). En plus des éléments décrits précédemment, la pharmacie (préparation, suivi) et les services connexes de nutrition ont été ajoutés aux coûts de base. L'évaluation des actes de diagnostic et les procédures médicales étaient basées sur l'unité de valeur relative et le coût par unité d'estimation du CA du département pour chaque établissement.

2.2.8. Honoraires des médecins

Les honoraires des médecins ont été estimés à partir des données provinciales de facturation médicale de la RAMQ, qui couvrent tous les actes médicaux réalisés par les médecins. Toutefois, il se peut que certains psychiatres soient payés à même le budget de leur centre hospitalier sous forme de salaire, ce

qui n'apparaît pas dans la banque de données administratives. Pour faire face à ce problème, les frais moyens chargés à la RAMQ par les psychiatres ont été appliqués pour chaque consultation ambulatoire dans les centres d'étude. Cet ajustement permet d'éviter la présence de biais d'information et de sous-estimation du coût des services psychiatriques.

2.2.9. Consommation de médicaments dans le contexte ambulatoire

Le registre des services pharmaceutiques de la RAMQ contient des informations concernant tous les médicaments délivrés pour chaque patient (sous le numéro d'assurance maladie). L'information comprend la date de dispensation, la dose, la quantité totale et le coût. Compte tenu du degré de sévérité de la maladie de notre cohorte de patients, nous supposons que tous les patients ont été couverts par le régime général d'assurance médicaments (RGAM) du Québec.

2.2.10. Logement

Les coûts liés au logement ont été déterminés par des entrevues et évalués sur une base quotidienne. L'évaluation a été basée sur les paiements quotidiens du ministère de la Santé alloués pour les résidences particulières (c.-à-d. appartements supervisés, foyers d'accueil, maisons de chambres et maisons de transition) (MSSS, 2001). Les patients adultes ont également contribué, à partir de leur prestation mensuelle d'aide sociale, aux frais de logement. Ces paiements ont été considérés pour la perspective du patient.

2.2.11. Analyse des données

Pour évaluer l'effet de la participation au Programme IPT sur les coûts en soins de santé et les coûts pour le patient, l'objectif de l'analyse économique était de comparer, chez la population à l'étude, les coûts engendrés durant l'année qui a précédé leur participation au programme IPT à ceux de l'année de leur participation à ce même programme. Les différences statistiques sur le plan de l'utilisation des services de santé et des coûts ont été évaluées à l'aide de tests des rangs signés de Wilcoxon pour les données non paramétriques (Armitage *et al.*, 2002), tenant ainsi compte du fait que les données de coûts ne présentaient pas une distribution normale. Les données ont été analysées en utilisant le logiciel statistique SPSS version 10.0 (SPSS Inc, Chicago, IL).

2.3. Résultats

2.3.1. Programme IPT

En fonction du nombre d'heures (en moyenne trois journées de huit heures) accordées à la formation et le coût moyen horaire des ergothérapeutes, le coût de la formation des professionnels en santé mentale pour le programme IPT a été estimée à 2347 $ (IC 95 % : 2260 $ – 2434 $) par professionnel. Les coûts annuels reliés au suivi du programme IPT, incluant principalement les honoraires des professionnels en santé mentale, sont de l'ordre de 1350 $ (IC 95 % : 1151 $ – 1549 $) par patient pour un total de 52 visites en moyenne.

2.3.2. Coûts ambulatoires et hospitaliers

Un résumé des coûts des services de santé offerts au cours du programme IPT, comparativement à ceux de l'année précédant son introduction, est présenté au tableau 11.3. Les résultats montrent que lors de la participation au programme IPT, il y a une diminution significative des coûts associés aux visites ambulatoires avec un psychiatre (618 $ contre 744 $, $p < 0{,}05$) et aux hospitalisations (4613 $ contre 6827 $, $p < 0{,}05$) comparativement à ceux de l'année précédant le programme IPT. Aucune différence significative n'a été observée entre les deux périodes pour le nombre de visites ambulatoires auprès de professionnels en santé mentale (excluant les visites effectuées dans le cadre du programme IPT) et avec d'autres médecins, pour le nombre et les coûts de visites à l'urgence et pour les coûts associés aux médicaments délivrés.

En limitant l'analyse aux personnes qui avaient participé à toutes les sessions du programme IPT ($n = 55$), une diminution significative a été observée au cours de l'IPT par rapport à l'année précédente quant au nombre d'admissions (< 1 par participant), à la durée du séjour (29 contre 11 jours) et aux coûts moyens (7127 $ [IC 95 % : 3169 $ – 11 085 $] contre 2809 $ [IC 95 % : 36 $ – 5655 $]).

2.3.3. Frais de logement

Dans l'année précédant le début de l'IPT, 50 % des participants vivaient dans une maison ou un appartement. Lorsque l'on examine les coûts en matière de logement, il n'y avait pas de différence significative dans le coût moyen par participant au cours de l'IPT (11 276 $), comparativement à celui de l'année précédant sa mise en œuvre (12 983 $).

Tableau 11.3.
Utilisation des ressources du système de santé (médiane) et coûts médicaux moyens par participant (n = 87)

	Pré-IPT	Durant l'IPT
Consultations ambulatoires avec psychiatre :		
N[bre] visites§	20 (5)	17 (12)
Coûts§	744 $ (529 $)	618 $ (445 $)
Frais d'actes médicaux§	531 $ (470 $)	363 $ (243 $)
Consultations ambulatoires avec d'autres professionnels en santé mentale* :		
N[bre] visites	37 (28)	42 (24)
Coûts	2529 $ (1179 $)	2760 $ (1271 $)
Consultations ambulatoires avec d'autres médecins ou spécialistes :		
N[bre] visites	4 (1)	3 (0)
Coûts	160 $ (0)	130 $ (0)
Frais d'actes médicaux	126 $ (0)	105 $ (0)
Département d'urgence		
N[bre] visites§	1 (1)	< 1 (0)
Durée de séjour (jours)	2 (1)	2 (0)
Coûts	431 $ (99 $)	318 $ (0)
Hospitalisations		
N[bre] admissions§	1 (0)	< 1 (0)
Durée de séjour (jours)§	27 (0)	18(0)
Coûts§	6827 $ (0)	4613 $ (0)
Actes diagnostic§	104 $ (0)	53 $ (0)
Frais d'actes médicaux§	406 $ (0)	162 $ (0)
Médicaments	3616 $ (2460 $)	4028 $ (2756 $)

* Les professionnels en santé mentale incluent ergothérapeutes, travailleurs sociaux, psychologues et infirmiers.
§ Différence significative (p < 0,05) selon les tests des rangs signés de Wilcoxon.

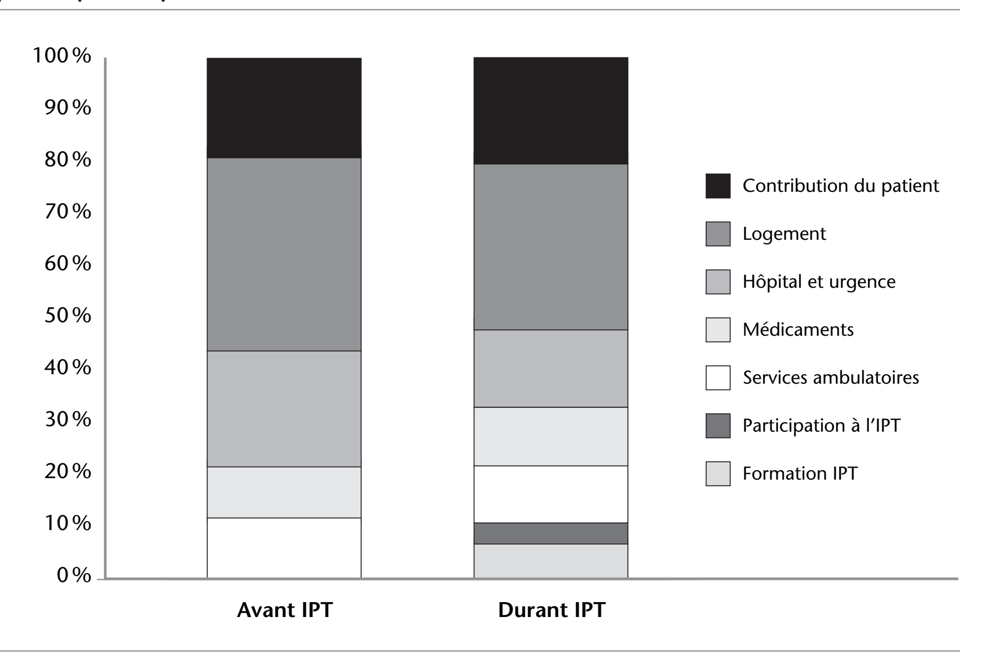
2.3.4. Coûts directs non médicaux (perspective du patient)

Le temps moyen pour la participation à l'IPT était de 82 heures, et les coûts moyens reliés au transport et à la participation au programme étaient de 238 $. Outre les coûts supplémentaires liés à la participation au programme, il n'y avait pas de différence significative en matière de temps passé à

rechercher des soins médicaux ambulatoires (55 heures comparativement à 58 heures) et de coûts liés au transport (263 $ comparativement à 267 $) entre les périodes avant et pendant l'IPT. Le montant alloué au logement était en moyenne 7032 $ par participant annuellement.

En résumé, le coût total moyen des soins de santé a été estimé à 26 750 $ par participant dans l'année précédant l'IPT et de 29 830 $ par participant lors de leur implication dans le programme IPT. Les coûts moyens supportés par le participant étaient respectivement de 7295 $ et de 7537 $ pour les périodes avant et pendant l'IPT. Un résumé de la contribution relative de chaque type de ressources utilisées est présenté à la figure 11.1.

Figure 11.1.
Sommaire de l'utilisation des services de soins de santé et des coûts supportés par le participant



En résumé, nos résultats montrent que le programme IPT a engendré 10 % des coûts observés au cours de la période de suivi. Les coûts liés au logement représentaient respectivement 33 % et 35 % des coûts avant l'IPT et pendant l'IPT. Les hospitalisations de courte durée et les visites à l'urgence au cours de l'IPT représentaient 14 % des coûts comparativement à 22 % dans l'année qui le précédait.

Cette analyse possède toutefois certaines limites. Premièrement, elle est basée sur un petit nombre de participants et une courte période de suivi ; un recrutement plus conséquent de personnes et un suivi plus long auraient permis de produire des estimations plus précises pour chaque type de clinique de psychiatrie. De plus, certains coûts associés à l'IPT n'ont pu être traités dans notre étude pour des raisons de faisabilité, dont les coûts indirects associés à la participation au programme IPT (p. ex., coûts reliés à l'absentéisme et au présentéisme au travail, coûts supportés pour la famille et les proches). Selon certaines études antérieures ayant évalué le coût de la schizophrénie, les coûts indirects peuvent atteindre 50 % des coûts totaux (Goeree *et al.*, 1999 ; Knapp, 1997). Ainsi, nos résultats peuvent sous-estimer la réduction des coûts non médicaux supportés par le patient et sa famille.

Cette analyse descriptive des coûts permet de conclure que le programme IPT engendre des coûts ponctuels fixes pour la formation des professionnels et des coûts externes supplémentaires pendant toute la durée des modules réalisés par les professionnels en santé mentale. Mis à part les coûts de formation, les sujets participant au programme IPT ont engendré des coûts moindres pour le système de santé au cours de l'IPT que l'année précédant son initiation. Ainsi, en plus d'avoir été démontré comme étant une intervention efficace (Brenner *et al.*, 1992 ; Mueller *et al.*, 2003 ; Briand *et al.*, 2006), le programme IPT possède des coûts inférieurs à tout autre type de soins de santé ; il est associé à des coûts minimes pour les personnes y participant et il peut avoir un effet positif de réduction des hospitalisations et d'autres types de soins ambulatoires chez les personnes aux prises avec une schizophrénie.

CONCLUSION

Bien que l'analyse des coûts soit très importante pour informer sur les composantes économiques d'une intervention, elle ne permet toutefois pas d'effectuer des recommandations pour l'allocation des ressources de la santé. Des évaluations économiques complètes seraient nécessaires afin d'évaluer le coût-efficacité d'une intervention (comme illustré avec le programme IPT) comparativement à d'autres interventions (p. ex., thérapies psychosociales ou d'autres traitements). De plus, cette analyse illustre bien certaines complexités de l'évaluation économique, dont la présence de plusieurs sources de données, la perte de suivi d'un nombre considérable de participants et un faible taux de participation. Un faible taux de participation et la perte de suivi sont toujours des préoccupations importantes pour la validité d'études épidémiologiques et économiques, notamment si les personnes n'ayant pas terminé le traitement ont des caractéristiques

différentes (p. ex., état de santé, utilisation des services de santé, réponse au traitement) de celles qui l'ont terminé. Pour cette raison, un maximum d'efforts est nécessaire pour minimiser la perte de suivi et obtenir un maximum d'informations sur ces sujets afin de bien analyser et de bien interpréter leur effet sur les résultats.

Un autre défi important rencontré lors de l'analyse des coûts (comme cela a été le cas pour l'IPT) est que les données de coûts peuvent provenir directement des différents centres hospitaliers inclus dans l'étude, au lieu d'être des données de coûts unitaires régionales ou nationales. Cette situation peut faire en sorte que les résultats perdent un peu de précision, mais cela permet la généralisation des résultats et la comparaison avec d'autres régions, et l'analyse de la tendance dans le temps. Ainsi, avec la présence d'une multitude de sources de données et l'utilisation de techniques de modélisation et de tests statistiques complexes et non intuitifs pour les preneurs de décisions, la règle d'or de l'évaluation économique est de démontrer une grande transparence dans l'analyse des données et dans l'interprétation des résultats, des limites de l'étude et de la transférabilité des résultats au contexte clinique.

Annexe 11.1.
Adaptation française de la grille d'évaluation de la qualité d'une évaluation économique de Drummond *et al.* (2005)

CRITÈRES D'ÉVALUATION DE LA QUALITÉ	OUI	NON	S. O.
1. A-t-on posé une question précise, à laquelle on puisse répondre ? ▪ L'étude prend en compte à la fois les coûts et les résultats des interventions. ▪ L'étude compare les différentes options. ▪ Un point de vue précis est adopté, et l'étude se place dans un contexte décisionnel particulier (*Perspective de l'étude*).	☐	☐	☐
2. Les options concurrentes sont-elles décrites de manière détaillée et complète ? ▪ Aucune option importante n'a été omise. ▪ L'option « ne rien faire » est envisagée et étudiée si pertinente. ▪ Les éléments descriptifs des options sont présentés (fréquence, population d'analyse, schéma de l'intervention, etc.).	☐	☐	☐
3. L'efficacité des interventions en pratique réelle est-elle établie ? ▪ L'efficacité est établie par un essai clinique randomisé et contrôlé, dont le protocole correspond à ce qui adviendrait en pratique courante. ▪ L'efficacité est établie par une synthèse d'études cliniques de bonne qualité méthodologique. ▪ L'efficacité est établie par des données d'observation ou des hypothèses avec une analyse des biais sur les conclusions.	☐	☐	☐
4. Les coûts et les effets sur la santé les plus importants de chaque option sont-ils déterminés ? ▪ Les différents points de vue pertinents sont examinés, tant pour les coûts que pour les effets sur la santé. ▪ Aucun effet sur la santé important n'est omis. Si un effet important n'est pas examiné, ce choix est justifié. ▪ Aucun coût important n'est omis. Si un élément de coût important n'est pas examiné, ce choix est justifié.	☐	☐	☐
5. Les coûts et les effets sur la santé sont-ils mesurés avec les unités appropriées (année de vie, événement évité, score de préférence, etc.) ? ▪ Les sources d'information sont clairement décrites et la source la plus pertinente est privilégiée. ▪ Y a-t-il des données omises ou difficilement mesurables ? Quel est l'effet potentiel sur les résultats ?	☐	☐	☐

(*suite*)

CRITÈRES D'ÉVALUATION DE LA QUALITÉ	OUI	NON	S. O.
6. Les coûts et les effets sur la santé ont-ils été mesurés avec rigueur ? ▪ Toutes les sources d'information sont clairement indiquées. ▪ Est-ce que les valeurs marchandes ont été employées pour déterminer le changement sur le plan des ressources (p. ex., perte de productivité) ? ▪ Si certaines valeurs marchandes étaient absentes, est-ce que des ajustements ont été effectués pour estimer ces dernières correctement ? ▪ Est-ce que la valorisation des conséquences a été effectuée de manière appropriée et en fonction du type d'étude (coût-efficacité, coût-bénéfice, coût-utilité) ?	☐	☐	☐
7. Les coûts et les résultats de santé futurs sont-ils ajustés en fonction du temps ? ▪ Les coûts et les résultats sont actualisés à un même taux. ▪ Le taux d'actualisation est connu et justifié.	☐	☐	☐
8. Est-ce qu'une analyse incrémentale des coûts et des conséquences a été effectuée ? ▪ Est-ce que les coûts additionnels (incrémentaux) de l'intervention, par comparaison avec son comparateur, ont été comparés aux effets additionnels de l'intervention ?	☐	☐	☐
9. A-t-on tenu compte de l'incertitude dans l'estimation des coûts et des résultats de santé ? ▪ Si les données le permettent, une analyse de sensibilité est-elle réalisée ? ▪ Les intervalles de distribution utilisés pour les analyses de sensibilité sont-ils justifiés ? ▪ L'incertitude attachée aux conclusions de l'évaluation économique est connue et discutée (intervalles de confiance, ellipses de confiance découlant des analyses de sensibilité, courbe d'acceptabilité).	☐	☐	☐
10. L'interprétation des conclusions de l'évaluation économique est-elle pertinente ? ▪ Une analyse différentielle des coûts et des résultats de santé des options concurrentes est réalisée et présentée. Est-ce fait selon les règles de l'art ? ▪ Les conclusions sont comparées de manière critique à celles d'autres études sur le même sujet. ▪ L'étude aborde la question de la généralisation des conclusions à d'autres contextes ou à d'autres groupes de patients. ▪ L'étude prend en compte d'autres facteurs entrant dans la décision (éthique, financement, organisation et mise en œuvre, etc.). ▪ L'étude discute de l'implémentation, comme la faisabilité d'adopter l'intervention.	☐	☐	☐

RÉFÉRENCES

American Psychiatric Association – APA (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4e éd., Washington, American Psychiatric Press.

Armitage, P., G. Berry et J.N.S. Matthews (2002). *Statistical Methods in Medical Research*, Oxford, Blackwell.

Brazier, J., J. Roberts et M. Deverill (2002). « The estimation of a preference-based measure of health from the SF-36 », *J Health Econ*, vol. 21, nº 2, p. 271-292.

Brenner, H.D., B. Hodel, R. Genner, V. Roder et P.W. Corrigan (1992). « Biological and cognitive vulnerability factors in schizophrenia : Implications for treatment », *Br J Psychiatry*, vol. 161, nº 18, p. 154-163.

Brenner, H.D., V. Roder, B. Hodel, N. Kienzle, D. Reed et R.P. Liberman (1994). *Integrated Psychological Therapy for Schizophrenic Patients*, Toronto, Hogrefe.

Briand, C., H.M. Vasiliadis, A. Lesage, P. Lalonde, E. Stip, L. Nicole *et al.* (2006). « Including Integrated Psychological Treatment (IPT) as part of standard medical therapy for patients with schizophrenia : Clinical outcomes », *J NervMent Dis*, vol. 194, nº 7, p. 463-470.

Brousselle, A., F. Champagne, A.P. Contandriopoulos et Z. Hartz (2009). *L'évaluation : concepts et méthodes*, Montréal, Les Presses de l'Université de Montréal.

Canadian Agency for Drug and Technologies in Health – CADTH (2006). *Guidelines for the Economic Evaluation of Health Technologies : Canada*, 3e éd., Ottawa, Canadian Agency for Drugs and Technologies in Health.

Canadian Coordinating Office for Health Technology Assessment – CCOHTA (1996). *A Guidance Document for the Costing Process*, version 1.0, Ottawa, Canadian Coordinating Office for Health Technology Assessment.

Contandriopoulos, A.P. (2005). « Les enjeux éthiques associés à la transformation des systèmes de soins », *Éthique publique*, vol. 5, nº 1, p. 42-57.

Crochard-Lacour, A. et J. LeLorier (2000). *Introduction à la pharmacoéconomie*, Montréal, Les Presses de l'Université de Montréal.

Drummond, M.F., M.J. Sculpher, M.J. Torrance, B.J. O'Brien et G.L. Stoddart (2005). *Methods for the Economic Evaluation of Health Care Programmes*, 3e éd., Oxford, Oxford University Press.

Evans, R.G. (1997). « Going for the gold : The redistributive agenda behind market-based Health Care Reform », *Journal of Health Politics, Policy and Law Offprint*, vol. 22, nº 2, p. 427-465.

Feeny, D., W. Furlong, M. Boyle et G.W. Torrance (1995). « Multi-attribute health status classification systems. Health Utilities Index », *Pharmacoeconomics*, vol. 7, nº 6, p. 490-502.

Goeree, R., B.J. O'Brien, P. Goering, G. Blackhouse, K. Agro, A. Rhodes *et al.* (1999). « The economic burden of schizophrenia in Canada », *Can J Psychiatry*, vol. 44, nº 5, p. 464-472.

GOEREE, R., F. FARAHATI, N. BURKE, G. BLACKHOUSE, D. O'REILLY, J. PYNE *et al.* (2005). « The economic burden of schizophrenia in Canada in 2004 », *Curr Med Res Opin*, vol. 21, n° 12, p. 2017-2028.

GROSSE, S.D. (2008). « Assessing cost-effectiveness in healthcare : History of the $50 000 per QALY threshold », *Expert Rev Pharmacoecon Outcomes Res*, vol. 8, n° 2, p. 165-178.

INSTITUTE OF HEALTH ECONOMICS – IHE (2000). *A National List of Provincial Costs for Health Care : Canada 1997/1998*, <http://www.ihe.ca>.

KESSLER, R.C., P.R. BARKER, L.J. COLPE *et al.* (2003). « Screening for serious mental illness in the general population », *Arch Gen Psychiatry*, vol. 60, n° 2, p. 184-189.

KNAPP, M. (1997). « Cost of schizophrenia », *Br J Psychiatry*, vol. 171, p. 509-518.

MANNING, W.G. et J. MULLAHY (2001). « Estimating log models : To transform or not to transform ? », *Journal of Health Economics*, vol. 20, n° 4, p. 461-494.

MINISTÈRE DE LA SANTÉ et DES SERVICES SOCIAUX – MSSS (2001). *Base de données APR_DRG 2000-2001*, version finale (AS-471), Québec, MSSS.

MUELLER, D.R., V. RODER et H.D. BRENNER (2003). « Effectiveness of the Integrated Psychological Treatment (IPT) for schizophrenia patients : A meta-analysis covering 25 years of research », *Schizophr Res*, vol. 60, p. 326.

OWENS, D.K. (1998). « Interpretation of cost-effectiveness analyses », *J Gen Intern Med*, vol. 13, n° 10, p. 716-717.

PATEL, V. (2000). « The need for treatment evidence for common mental disorders in developing countries », *Psychological Medicine*, vol. 30, n° 4, p. 743-746.

PHELPS, C.E. (2003). *Health Economics*, 3e éd., Boston, Addison-Wesley.

PRINCE, M., R. STEWART, T. FORD et M. HOTOPF (2003). *Practical Psychiatric Epidemiology*, Oxford, Oxford University Press.

SMITH, D.H. et H. GRAVELLE (2000). *The Practice of Discounting Economic Evaluation of Health Care Interventions*, CHE Technical Paper Series 19, Center for Health Economics, The University of York, <http://www.york.ac.uk/che/pdf/tp19.pdf>.

SONNENBERG, F.A. et J.R. BECK (1993). « Markov models in medical decision making : A practical guide », *Med Decis Making*, vol. 13, n° 4, p. 322-338.

TARRICONE, R. (2006). « Cost-of-illness analysis. What room in health economics ? », *Health Policy*, vol. 77, n° 1, p. 51-63.

THE EUROQOL GROUP (1990). « EuroQol – A new facility for the measurement of health-related quality of life », *Health Policy*, vol. 16, n° 3, p. 199-208.

VASILIADIS, H.M., C. BRIAND, A. LESAGE *et al.* (2006). « Health care resource use and cost associated with integrated psychological treatment », *J Ment Health Policy Econ*, vol. 9, n° 4, p. 201-207.

WOLFF, N., T.W. HELMINIAK et R.J. DIAMOND (1995) « Estimated societal costs of assertive community mental health care », *PsychiatrServ*, vol. 46, n° 9, p. 898-906.

WORLD HEALTH ORGANIZATION – WHO (2003). *The Mental Health Context. Mental Health Policy and Service Guidance Package*, <http//www.who.int>.

CHAPITRE 12

LE GROUPE DE DISCUSSION FOCALISÉE
Application pour recueillir des informations sur le fonctionnement au quotidien des personnes avec un trouble de la personnalité limite

Julie Desrosiers
Nadine Larivière

FORCES

- Il favorise l'expression et l'émergence de différents points de vue.
- Il permet l'inclusion de participants moins scolarisés ou présentant un niveau de littératie faible.
- Son appropriation est aisée lorsque l'animateur possède de l'expérience en animation de groupes.

LIMITES

- Il demande un temps de préparation et d'analyse considérable.
- Il y a un risque de domination de certains participants dans le groupe.
- Il peut être difficile pour l'animateur-clinicien d'éviter d'intervenir lorsque des participants présentent une souffrance clinique.

Le groupe de discussion focalisée (aussi appelé *focus group* ou groupe focalisé) est une méthode d'enquête qualitative mise au point dans les années 1940 par les sociologues Paul Lazarsfeld et Robert Merton dans le domaine du marketing (Baribeau, Luckerhoff et Guillemette, 2010). Elle consiste à rassembler un petit groupe d'individus homogène appartenant à une population précise, afin de provoquer des échanges et des discussions autour de thèmes prédéfinis. Le principe essentiel consiste en ce que le chercheur utilise explicitement l'interaction entre les participants, à la fois comme moyen de collecte de données et comme point de focalisation dans l'analyse (Baribeau et Germain, 2010; Kitzinger, 1995; Gronkjaer *et al.*, 2011; Puchta et Potter, 2004). Le groupe est animé par un modérateur familier avec la méthodologie et apte à exploiter la dynamique propre au rassemblement en groupe. Le but est d'en arriver, pour les chercheurs, à une compréhension riche et approfondie de l'expérience et des croyances des participants sur le thème d'intérêt.

La méthode consiste plus précisément à interroger des petits groupes de personnes (environ de cinq à douze participants) répondant à des critères d'homogénéité par rapport à leur expérience du sujet de discussion afin de susciter un échange ouvert. La discussion est menée à partir d'une grille d'entrevue définissant les thèmes de l'étude. Par la suite, la synthèse et l'analyse du contenu sont réalisées afin de relever les principaux messages clés émis par les participants (Krueger, 1994).

Selon Morgan (1998), spécialiste de cette méthode de recherche, il s'agit essentiellement d'une manière d'écouter les personnes et d'apprendre d'elles. En regroupant des personnes différentes mais qui ont un rapport commun avec le thème de recherche, on vise à générer des discussions permettant de comprendre les similitudes et les divergences en ce qui a trait aux pensées, aux points de vue et aux émotions des participants. D'elles-mêmes, les personnes du groupe se comparent afin de saisir ce qui les relie et ce qui les différencie: le rôle de l'animateur est de les guider avec des questions afin d'obtenir un éventail de réponses, puis à analyser les données afin d'en faire ressortir ce qui a de l'importance pour ces personnes, par rapport à ce thème-là, dans ce contexte particulier.

Dans la première partie de ce chapitre, il sera question des différentes applications de cette méthode ainsi que de leurs critères de qualité, des procédures de mise en œuvre et d'analyse, et, pour terminer, des distinctions avec d'autres méthodes d'enquête. La seconde partie portera sur une illustration pratique de l'utilisation des groupes de discussion focalisée auprès de personnes avec un trouble de la personnalité limite pour mieux comprendre le concept du fonctionnement au quotidien de cette clientèle.

1. PRÉSENTATION DE LA MÉTHODE DES GROUPES DE DISCUSSION FOCALISÉE

1.1. Dans quels contextes cette méthode est-elle utile ?

Comme toutes les méthodes qualitatives, la méthode des groupes de discussion focalisée sert particulièrement à l'exploration et à la découverte de sujets peu ou mal connus, ainsi qu'à avoir accès aux points de vue des personnes concernées par le thème (Kitzinger, 1995). Le groupe de discussion focalisée est également une bonne méthode pour cibler les principales problématiques à explorer dans le cadre d'une enquête plus approfondie. De même, en ce qui concerne les comportements complexes et multifactoriels, la richesse d'une discussion de groupe permet d'en explorer les motivations, le comment et le pourquoi, de même que les divergences (Kitzinger, 1995 ; Morgan, 1998). Finalement, cette méthode peut être utilisée quand il y a une divergence (*gap*) entre des groupes de personnes (p. ex., entre thérapeutes et utilisateurs de service, ou encore entre chercheurs et cliniciens). Les groupes de discussion focalisée peuvent contribuer à mettre cette différence en lumière et à la comprendre, ou encore aider à la diminuer, en examinant en profondeur les divers points de vue (Kitzinger, 1995).

1.2. Quels sont les buts de l'utilisation des groupes de discussion focalisée ?

Dans la recherche en contexte de réadaptation en santé mentale et dans d'autres disciplines, les groupes de discussion focalisée peuvent être utilisés dans différents types d'études et à divers moments d'un projet de recherche (Dawson et Manderson, 1993).

Pendant la phase de découverte de l'origine du problème, les groupes de discussion focalisée peuvent servir à générer les questions de recherche, à mettre en lumière les besoins d'une population particulière ou à déterminer les parties prenantes lors de la phase de conceptualisation d'un projet. Les groupes de discussion focalisée constituent une méthode de choix pour l'étude de thèmes sur lesquels il y a peu d'écrits (Krueger et Casey, 2009). Par exemple, dans le cadre d'un projet de recherche fictif (utilisé ci-après à des fins didactiques) s'intéressant à la problématique de l'itinérance au centre-ville de Montréal, les professionnels de la santé impliqués dans le suivi des personnes itinérantes pourraient être interviewés lors de groupes focalisés afin de connaître leur opinion sur les services lacunaires pour cette clientèle.

Pendant la planification d'un projet, le groupe de discussion focalisée peut aussi être utilisé afin de décider du devis de la recherche et des moyens prévus pour atteindre le but fixé, de tester la faisabilité d'un projet, de procéder à certains choix méthodologiques ou de donner des balises

procédurales au projet de recherche. La consultation des différentes parties prenantes et l'implication de celles-ci dans le processus décisionnel favorisera l'application des connaissances produites. Pour reprendre le même exemple fictif vu plus haut, le même groupe de professionnels pourrait être questionné sur leur désir de collaborer à un projet de recherche sur l'intégration des meilleures pratiques et sur leur opinion quant à l'implication des usagers dans l'élaboration du projet.

Lors de la mise en œuvre d'un projet, le groupe de discussion focalisée peut être utilisé afin de produire des connaissances sur un thème encore peu connu. Cette méthode peut permettre de faire émerger un corpus thématique sur un sujet donné et mener à la conceptualisation de la problématique centrale du projet. Suivant le même exemple fictif, les chercheurs pourraient rencontrer des personnes itinérantes lors de groupes focalisés afin de s'intéresser aux motifs invoqués par ceux-ci pour vivre dans la rue.

Lors de l'analyse de résultats ou dans le cadre de l'évaluation de programmes ou de services, le groupe de discussion focalisée peut servir d'outil d'évaluation. Par exemple, le groupe de discussion focalisée peut permettre de recueillir les avis sur le déroulement d'un projet pour en connaître les retombées bénéfiques et les effets indésirables (Krueger et Casey, 2009). La désignation des leviers et des obstacles par les parties prenantes peut aider au changement de pratiques dans les organisations. Pour reprendre le même exemple fictif, l'équipe de recherche pourrait rencontrer en groupe des personnes itinérantes et des professionnels à la suite de l'implantation d'une ressource d'hébergement au centre-ville afin de recueillir leurs avis (voir le chapitre 23 de cet ouvrage).

1.3. Caractéristiques de la méthode

Le groupe de discussion focalisée est une méthode de recherche qualitative qui laisse place à une grande flexibilité, du point de vue tant de la structure, de l'échantillonnage que de l'analyse. Il est donc primordial de bien saisir les divers critères de qualité pouvant assurer des résultats fiables et permettant de justifier les choix méthodologiques, ainsi que le raisonnement ayant mené aux conclusions présentées. Les notions générales concernant la réalisation d'un guide d'entrevue (canevas du groupe de discussion) seront aussi présentées plus loin (voir section 1.5.1).

1.4. Critères de qualité

La méthode des groupes de discussion focalisée doit être appuyée par un canevas de raisonnement permettant au chercheur de justifier les diverses étapes, choix et conclusions de son projet, car aucune balise numérique ou procédurale fixe n'existe en définitive (p. ex., sur le nombre de rencontres). Les raisons qui motivent le choix du dispositif et la sélection des participants ont des répercussions sur le type de données à recueillir et sur la façon de les analyser (Baribeau, 2009). Les critères de qualité à considérer incluent la composition des groupes, le nombre de groupes nécessaires et de séances requises, ainsi que la façon d'analyser les données obtenues.

1.4.1. Composition des groupes

Plusieurs auteurs (Kitzinger, Markova et Kalampalikis, 2004; Morgan, 1998) s'entendent pour dire que le format type des groupes de discussion focalisée comprend cinq à huit participants. D'autres auteurs mentionnent plutôt une taille de groupe variant de six à douze participants (Geoffrion, 2009; Touré, 2010). Ce nombre est une moyenne, fondé sur l'expérience pratique, plutôt que sur des considérations théoriques, pour que tous aient la chance de s'exprimer au cours d'une séance. Un groupe incluant un nombre plus élevé de personnes (p. ex., plus de six personnes) implique nécessairement moins de temps de parole par personne et rend plus difficile la création d'une ambiance accueillante, mais il permet d'explorer plus largement un thème. Un groupe plus petit (p. ex., cinq personnes et moins) implique une moins grande diversité d'opinions, moins de comparaisons possibles spontanément entre les participants, ainsi que plus de pression à participer pour chacun des membres (Krueger et Casey, 2009).

Dans un groupe de discussion focalisée, il n'est pas possible d'estimer de façon mathématique la représentativité statistique de l'échantillon. La composition de chacun des groupes doit être réfléchie et justifiée par l'équipe de recherche: dans le protocole et le rapport final, l'arrêt du choix des participants doit être présenté et raisonné, avec les limites inhérentes à la méthode de recrutement, aux ressources disponibles et au contexte local (Baribeau, 2009). Ici, une certaine variété est visée dans les critères d'inclusion et d'exclusion choisis: le spectre d'âge est couvert, les deux genres sont représentés de façon équilibrée, diverses origines socioéconomiques ou culturelles sont représentées, etc. La considération d'une ou plusieurs caractéristiques de sélection implique l'hypothèse d'un effet quelconque de celle-ci sur les résultats. De plus, l'échantillon étant fréquemment divisé en plusieurs groupes (afin de respecter une taille pratique), les répercussions de

cette distribution hétérogène entre les divers groupes doivent être prises en compte pour qualifier les échanges au sein d'un même groupe, tout comme les différences entre les divers groupes (Morgan, 1998).

Bien qu'une variété soit nécessaire afin d'assurer un degré de représentativité à l'étude, l'homogénéité des participants de chacun des groupes est primordiale (Kitzinger *et al.*, 2004). En effet, comme il a été présenté plus haut, la nature sociale du groupe de discussion implique que l'ambiance et le confort des participants, entre eux et avec les animateurs, influent sur les échanges réalisés (Richarson et Rabiee, 2001). Dans ce cas, il faut s'assurer que les participants présentent une relation commune au thème de discussion et que les caractéristiques de certains participants n'inhibent pas la capacité à s'exprimer de certains autres. Si le thème de la recherche implique des expériences suffisamment différentes en fonction de caractéristiques démographiques ou autres, qui peuvent changer les expériences vécues au point de rendre des participants étrangers les uns aux autres ou mal à l'aise de parler de leur réalité, il peut devenir nécessaire de segmenter l'échantillon selon ces mêmes caractéristiques. Ceci multipliera le nombre de groupes, mais assurera une plus grande qualité des échanges.

1.4.2. Nombre de groupes et de séances

Le nombre de groupes de même que le nombre de séances pour chacun de ces groupes ne peuvent être déterminés d'avance, puisque cela dépend du thème de recherche et des participants recrutés. Le critère central ici est la saturation théorique : le nombre de groupes et de séances nécessaires est confirmé lorsqu'aucun nouvel élément utile à l'analyse n'apparaît lors d'une nouvelle discussion en groupe (Krueger, 1994). La flexibilité du protocole implique la possibilité de s'adapter au déroulement des groupes en cours. En général, au moins trois groupes différents sont nécessaires par thème, par segment de population ou par étape de la recherche, afin d'assurer une comparabilité. Ici encore, la justification raisonnée du choix du nombre de groupes et de séances est ce qui détermine ce choix. Plusieurs groupes sont souvent utiles afin de s'assurer que les sujets soulevés lors du groupe de discussion et les différentes expériences relevées ne sont pas le fruit du hasard de la composition d'un groupe en particulier, ceci afin d'obtenir des données variées, représentatives et comparables (Kitzinger *et al.*, 2004 ; Morgan, 1998).

1.4.3. Analyse des données recueillies

Les données recueillies à l'aide de cette méthode sont des conversations. En effet, les paroles recueillies ne peuvent être prises telles quelles, mais doivent être analysées en les qualifiant à l'aide du contexte dans lequel elles ont émergées, soit des dynamiques sociales d'un groupe de discussion, lesquelles sont dépendantes de la composition de celui-ci (Gronkjaer *et al.*, 2011). Les idées exprimées doivent être qualifiées à l'aide des éléments non verbaux les ayant accompagnées, mais aussi des dynamiques conversationnelles et argumentatives ayant présidé à leur émergence, de même qu'à l'aide des suites d'idées, recherches de consensus ou polarisations observées (Gronkjaer *et al.*, 2011 ; Kitzinger *et al.*, 2004).

1.5. Procédure

La réalisation d'un projet impliquant des groupes de discussion focalisée peut être organisée en quatre étapes, soit : 1) la planification ; 2) le recrutement des participants ; 3) la conduite des groupes ; et 4) l'analyse et la rédaction du rapport (Morgan, 1998). Ces étapes seront décrites en détail ci-dessous, puis le groupe de discussion sera comparé à d'autres méthodes de recherche, afin de mieux en saisir les caractéristiques particulières et d'éviter de l'utiliser pour des fins qui ne sont pas appropriées.

1.5.1. Planification

L'étape préparatoire peut être la plus à risque de négligence (Morgan, 1998), puisque la réalisation d'une discussion de groupe peut paraître simple et directe. Pourtant, la minutie et la portée de la préparation influencent grandement la qualité de l'étude. En effet, afin de bien orienter les discussions des groupes, il est essentiel que le sujet d'étude ait été clarifié et que les connaissances aient été explorées par une recension des écrits et l'exploration de sujets connexes. Ceci guidera la préparation des questions de discussion, qui doivent être assez précises pour permettre une meilleure délimitation du sujet, mais assez ouvertes et flexibles pour permettre de suivre le groupe où la discussion le mène, tout en le gardant sur les rails du thème d'intérêt. Selon Krueger (1998), la qualité des données recueillies dépend principalement des questions posées, avant même la manière d'animer les groupes ou de recueillir les données. Au-delà de la préparation des questions, indiquant le contenu et l'orientation globale de la discussion, l'étape de planification doit s'intéresser aux ressources matérielles et financières disponibles pour le projet de recherche : les locaux, les rafraîchissements, le

matériel d'enregistrement, la publicité, les affichages, le dédommagement des participants, la taille de l'équipe de recherche et le temps disponible pour la transcription des conversations et l'analyse.

1.5.2. Recrutement

Le choix des participants aux groupes se fait habituellement sur un mode raisonné plutôt qu'aléatoire. En effet, puisque cette méthode de recherche implique des discussions de groupe, qui à leur tour nécessitent une certaine ambiance et un confort entre les participants, ainsi qu'un certain spectre de diversité d'expériences, le recrutement des sujets de recherche se fera en prenant en compte deux aspects principaux, soit la compatibilité et la représentativité (Morgan, 1998).

1.5.3. Conduite des groupes

La réalisation des groupes de discussion focalisée constitue bien évidemment le cœur de cette méthode de recherche : un petit groupe de personnes se retrouvent, durant une période variant entre 90 minutes et 2 heures, et souvent à quelques reprises, avec un animateur qui leur soumet des questions et modère la discussion. Afin d'assurer la qualité des résultats, voici quelques éléments à prendre en compte.

Animation

Le rôle d'animateur est central pour le confort des participants, l'animateur s'assurant que tous se sentent à l'aise et pris en compte, et que la parole est équitablement partagée. Il doit aussi tendre à ce que les participants échangent entre eux et non qu'ils discutent tous avec l'animateur (Kitzinger *et al.*, 2004). La conduite d'une discussion en groupe, visant l'écoute des expériences et des opinions des participants, nécessite une flexibilité conversationnelle, afin de permettre l'émergence de nouveaux sujets et de pensées spontanées (Jovchelovitch, 2004). L'animateur doit les prendre en compte et les suivre, afin de bien saisir les raisonnements des participants, tout en s'assurant que ceux-ci sentent que leur opinion est pertinente à l'exercice (Morgan, 1998). D'un autre côté, l'absence d'un cadre bien défini et bien planifié peut mener à un trop grand éparpillement et diminuer la pertinence des propos recueillis. Pour faciliter le travail d'animation, un guide d'entrevue se doit d'être élaboré de la manière la plus précise et exhaustive possible. Celui-ci vise à aider le maintien des échanges sur le thème d'étude tout en explorant les sujets amenés par les participants et en relançant la discussion lorsque les échanges se tarissent (Puchta et Potter, 2004).

Guide d'entrevue

Un guide d'entrevue consiste en un ensemble de questions ouvertes, qui appellent des réponses et une discussion descriptives. Il s'agit d'un outil d'orientation de la discussion, qui doit être conçu pour garantir un rythme régulier de discussion, tout en s'assurant qu'aucune des questions clés ne sera omise. Ces questions clés devraient être en un nombre ni trop important (menaçant de diminuer la richesse des échanges) ni trop limité (laissant peu de possibilité de relance). Plusieurs auteurs (Morgan, 2004; Smithson, 2000) situent ce nombre entre cinq et huit, pour un groupe de 90 minutes de discussion, en présence de cinq à huit participants. Le guide d'entrevue comprend les thèmes d'interrogation (questions clés), suivis d'une série de questions de clarification, qui visent à apprendre pourquoi, comment, où, qui, avec qui, à quelle fréquence, etc. (Krueger, 1998b). Il est habituellement utile de débuter par des questions préliminaires, qui aident les personnes présentes à se familiariser avec l'ambiance de la discussion avant d'entrer dans le vif du sujet. Ensuite, les objectifs centraux du projet de recherche sont abordés, après avoir été traduits en thèmes d'interrogation. Ces thèmes sont abordés dans un ordre réfléchi, du général vers le particulier, afin de mettre en place un contexte, par exemple du passé vers l'avenir et de la réponse spontanée vers la clarification. Le langage utilisé a aussi son importance: un langage simple, accessible pour les participants, lié à leur réalité, est encouragé. Dans le cas de thèmes émotifs ou liés à des comportements, les questions de type «Pourquoi?» sont à considérer avec circonspection, puisqu'elles amènent des réponses justificatives et rationnelles. Demander à quelqu'un de décrire ce qui l'a mené à un choix, plutôt que simplement pourquoi ce choix a été fait, donne un accès plus fidèle à la conduite réelle (Krueger, 1998b).

Coanimation

La présence d'un coanimateur est souhaitable, car ce dernier peut gérer le temps et noter les éléments de la discussion qui n'apparaîtront pas dans l'enregistrement: aspects non verbaux de l'animateur et des participants, dynamique du groupe, climat affectif, etc. Ce même coanimateur peut aussi donner de la rétroaction à l'animateur pour la planification des rencontres suivantes (Krueger, 1998b).

Ambiance

L'ambiance qui sert de cadre aux discussions nécessite peu de préparatifs, mais requiert assurément une attention particulière de la part des chercheurs (Morgan, 1998). À cet effet, le choix du lieu peut être particulièrement

sensible dans le cas du domaine de la santé mentale; les locaux auront fréquemment un lien avec les lieux de traitement ou de recherche psychiatriques: ces éléments créeront-ils un biais important dans la recherche? Par ailleurs, toutes les autres petites attentions rendant les personnes à l'aise sont aussi à prévoir: rafraîchissements, pauses, chaises confortables, calme, éclairage, etc.

Confidentialité, anonymat et sensibilité

Il est essentiel d'informer les participants du respect de leur anonymat lors de l'analyse des données, par l'utilisation de pseudonymes et de codes, connus uniquement du responsable du projet. Le respect de la confidentialité de la part des participants est aussi primordial. Si l'on veut assurer le confort des participants durant les discussions et le respect de leur sensibilité, l'information et les identités dévoilées ne devront pas sortir du cadre du groupe (Krueger, 1998b).

1.5.4. Analyse et rédaction du rapport

L'analyse des données recueillies est probablement l'étape la plus longue de ce type de méthode. En effet, il est nécessaire de transcrire tous les échanges puis de dépouiller les paroles pour en faire ressortir les thèmes, les convergences et les divergences, avant d'en tirer des conclusions. Une analyse de qualité présente les caractéristiques suivantes: elle est séquentielle, continue, systématique et vérifiable (Krueger et Casey, 2009). Brièvement, ceci signifie que l'analyse est effectuée au fur et à mesure jusqu'à l'atteinte de la saturation des données (séquentielle), qu'elle est faite de façon concomitante à la collecte de données et qu'elle se poursuit tout au long du processus (continue), qu'elle tient compte de toutes les données verbales et interactionnelles (systématique), et, enfin, qu'elle est rapportée avec suffisamment de clarté pour être reproductible (vérifiable).

Il est conseillé d'effectuer l'analyse en plusieurs étapes, et ce, dès la fin d'une séance de groupe: les premières impressions sont cruciales (Morgan, 2004). C'est à ce moment que les observations sur les dynamiques interactionnelles et les contenus verbaux sont les plus fraîches et les plus claires (Kitzinger, 1995). À la suite de la transcription, il est nécessaire d'effectuer une indexation ou codage, permettant de faire ressortir les thèmes et les mots clés répertoriés. En plus de l'identification du contenu des échanges, il importe d'appuyer ceux-ci à l'aide des aspects interactifs et conversationnels: censure, argumentation, formation d'un consensus, éléments non verbaux, etc. (Gronkjaer *et al.*, 2011; Kitzinger *et al.*, 2004). Le contexte dans lequel a eu lieu l'entrevue de groupe doit aussi être considéré: des éléments

tels que le climat, l'ambiance, les habiletés de l'animateur, les réactions des participants sont autant de facteurs devant être inclus dans l'analyse (Vicsek, 2007). Il est bien sûr essentiel de considérer les mots utilisés lors des groupes, ainsi que leur fréquence d'apparition (afin de demeurer fidèle aux idées des participants), mais le chercheur ne doit pas confondre la fréquence d'un thème et son importance. Il n'y a en effet pas de données objectives liant l'importance d'un thème à sa fréquence (Krueger, 1998b). Les formulations peuvent en effet varier entre les différentes personnes et l'importance d'un aspect pour la recherche peut différer de celle que les participants ressentent. Également, la tension liée à un sujet de discussion peut rendre l'élaboration verbale plus ardue, ce qui n'implique pas que cet aspect soit de moindre importance. Un énoncé émis par un participant peut aussi rapidement créer un consensus et ne pas nécessiter une longue discussion, ce qui n'implique pas qu'il soit secondaire.

L'analyse mène à des conclusions argumentées, formulées à l'aide de ce que les participants ont soulevé, en les liant à ce qui est déjà connu. Les questions clés utilisées serviront à organiser les données. L'analyse des données devraient se faire de façon concurrente à la collecte de données (Rabiee, 2004). En effet, l'élaboration d'hypothèses en cours de recherche, à l'aide des thèmes émergeant des discussions, permet d'orienter les discussions des groupes suivants en faisant ressortir de nouveaux thèmes à fouiller ou à creuser plus avant. Les conclusions présentées doivent montrer clairement le raisonnement suivi pour en arriver à la finalité des différentes étapes : planification (thèmes, questions clés, guide d'entrevue), recrutement (ces personnes-là, dans ces groupes, formés de cette façon), conduite (animation, reformulation, exploration, changements de sujets) et analyse (éléments conservés et rejetés, importance accordée aux différents thèmes et échanges).

1.6. Matériel et logiciels

La méthode des groupes de discussion focalisée est simple et directe, ce qui n'implique pas nécessairement qu'elle soit toujours abordable ou rapide, bien au contraire. En effet, les ressources matérielles nécessaires peuvent se révéler importantes, par exemple si la location de locaux devient nécessaire, ou si la mise sous contrat d'animateurs professionnels devient incontournable en raison d'un manque d'expérience dans ce domaine de la part des membres de l'équipe de recherche (Morgan, 1998). Le recrutement exigera de son côté des heures de travail et peut-être des coûts de publicité, ainsi que des frais de dédommagements, s'ils sont jugés nécessaires à la participation des sujets au projet de recherche. Également, de nombreuses

heures peuvent être nécessaires pour transcrire (car cette transcription est plus complexe que celle d'un entretien individuel), coder et analyser les données recueillies, ce qui peut représenter des coûts considérables. L'utilisation de logiciels (comme NVivo ou QDAMiner) n'est pas essentielle, mais peut faciliter l'analyse.

1.7. Distinctions avec d'autres méthodes apparentées

Quelques remarques sont proposées ici pour éviter de confondre les groupes de discussion focalisée avec d'autres méthodes. Tout d'abord, il importe de rappeler que le groupe de discussion focalisée n'est pas un lieu pour convaincre, enseigner ou organiser les participants ; il s'agit d'un espace créé pour les écouter et apprendre d'eux (Morgan, 1998). De plus, le groupe de discussion focalisée n'est pas une méthode de recherche appropriée si les chercheurs souhaitent pouvoir généraliser les résultats à une population statistique. Il s'agit plutôt de l'analyse d'une situation particulière, et son mode d'échantillonnage ne peut pas être statistiquement représentatif d'un groupe plus large (Krueger, 1998b). Le tableau 12.1 présente quelques différences avec d'autres méthodes de collecte de données apparentées, afin de mieux saisir l'utilité et la spécificité des groupes de discussion focalisée (Dalkey et Helmer, 1963 ; Krueger et Casey, 2009 ; Morgan, 1998 ; Powell et Single, 1996).

Dans le cadre de la recherche dans le domaine de la santé mentale, il est nécessaire de différencier le groupe de discussion focalisée du *groupe de thérapie*. En effet, les attentes de chacun ne sont pas les mêmes, ainsi que leur ambiance et leurs techniques d'animation. Bien sûr, l'expérience dans le domaine de la thérapie de groupe peut être aidante pour l'animateur, mais la confusion peut être dommageable pour les résultats comme pour les participants. Il faut s'assurer que les membres du groupe comprennent bien que même si les discussions portaient éventuellement sur des problèmes personnels vécus, le groupe de discussion focalisée n'a pas comme but premier de les aider à trouver des solutions à ces problèmes, même s'il peut amener une mise à distance et une meilleure compréhension souvent bénéfiques pour l'individu. La présence d'un coanimateur peut permettre d'éviter des débordements et assurer un soutien à une personne en détresse à la suite d'une discussion particulièrement éprouvante. L'équipe de recherche doit aussi réfléchir aux ressources et aux services vers lesquels les participants peuvent être dirigés en cas de besoin. La distribution générale et anonyme de ces informations, en début de séance, permet de cadrer le groupe et d'éviter la stigmatisation d'une personne en particulier en lien avec un besoin de soutien ou de service (Morgan, 1998).

Tableau 12.1.
Différences entre le groupe de discussion focalisée et des méthodes apparentées

Groupe de discussion focalisée	Sondage de groupe
Échantillonnage réfléchi.	Échantillonnage aléatoire ou stratifié.
Guide d'entrevue permettant une certaine flexibilité des thèmes abordés.	Questionnaire standardisé.
Groupe de discussion focalisée	**Méthode Delphi (voir le chapitre 13 de cet ouvrage)**
Méthode et analyse centrées sur les interactions entre les participants.	Aucune interaction entre les sujets afin de comparer les réponses individuelles.
Guide d'entrevue permettant une certaine flexibilité des thèmes abordés.	Questionnaire standardisé.
Conclusion de l'étude : une compréhension en profondeur de l'expérience des participants.	Conclusions de l'étude : la somme des opinions d'experts.
Groupe de discussion focalisée	**Observation en milieu naturel**
Situation artificielle créée afin d'avoir accès aux perceptions, aux croyances et aux attitudes.	Milieu naturel comme contexte d'observation des comportements.
Groupe de discussion focalisée	**Groupe nominal**
Méthode qui permet l'exploration d'un thème.	Méthode qui permet le classement ou la priorisation de sujets (p. ex., causes, solutions).
Discussions suivant le guide d'entrevue.	Alternance de phases de réflexions individuelles et de discussions de groupe.
Les discussions constituent l'essentiel du matériel de travail.	Utilisation d'une grille de réponse établie à l'avance et de questionnaires version papier.
Conclusion de l'étude : une compréhension en profondeur de l'expérience des participants.	Conclusions de l'étude : un ensemble d'avis indépendants menant à une opinion collective.

2. APPLICATION POUR RECUEILLIR DES INFORMATIONS SUR LE FONCTIONNEMENT AU QUOTIDIEN DES PERSONNES AVEC UN TROUBLE DE LA PERSONNALITÉ LIMITE

2.1. Présentation du projet

La clinique des troubles relationnels et de la personnalité de l'Institut universitaire en santé mentale de Montréal est une clinique spécialisée offrant des services de deuxième et troisième ligne dans le traitement des troubles de

la personnalité. L'équipe multidisciplinaire a décelé certaines lacunes sur le plan de l'évaluation, notamment en ce qui a trait à la mesure des effets du trouble de la personnalité limite (TPL) sur le fonctionnement quotidien. À l'heure actuelle et selon la connaissance que nous en avons, aucun outil de mesure du fonctionnement n'est bien adapté pour cette clientèle. L'objectif principal de ce projet est de concevoir et de valider un outil de mesure des effets fonctionnels et sociaux des personnes ayant un TPL. Des groupes de discussion focalisée réalisés en collaboration avec les personnes présentant un TPL constituent l'une des principales méthodes de collecte de données dans ce projet. Leur utilité est de déterminer les dimensions du fonctionnement à mesurer de même que de générer les items de l'outil de mesure. Dans les sections suivantes, nous nous concentrerons sur les buts et l'utilisation concrète des groupes de discussion focalisée auprès de personnes avec un TPL dans le développement d'un outil de mesure.

2.2. Buts de l'utilisation des groupes de discussion focalisée

Au cours d'une rencontre de consultation préliminaire, l'équipe de recherche et les professionnels de la santé ont décidé de procéder à l'élaboration d'un outil de mesure autoadministré. L'élaboration de mesures autorapportées par l'intermédiaire de groupes de discussion focalisée connaît présentement un essor dans les écrits scientifiques (Lasch *et al.*, 2010) puisque l'on reconnaît l'importance d'accorder une voix aux sujets dans la conception d'outils les concernant, mais aussi parce que le contenu des items gagne en validité et en crédibilité si les personnes avec des troubles mentaux reconnaissent l'authenticité des termes utilisés pour décrire leur réalité (voir le chapitre 24 de cet ouvrage). Des groupes de discussion focalisée ont donc été mis en œuvre dans ce projet dans le but de recueillir l'expérience des personnes vivant avec un TPL face à leur fonctionnement au quotidien et, ultérieurement, de contribuer à la création des items de l'outil.

Préalablement à la tenue des groupes de discussion focalisée avec les personnes ayant un TPL, un processus de consultation a été réalisé auprès de vingt professionnels de la santé œuvrant auprès de ces personnes. L'objectif était que ces professionnels déterminent les dimensions du fonctionnement les plus significativement touchées chez cette clientèle, à partir de la Classification internationale du fonctionnement, du handicap et de la santé de l'Organisation mondiale de la santé (OMS, 2010). Au total, les professionnels consultés ont choisi 17 dimensions (p. ex., effectuer la routine quotidienne et gérer le stress et autres exigences psychologiques), ainsi que 38 sous-dimensions du fonctionnement pouvant être affectées (p. ex., gérer

son propre niveau d'activité et assumer ses responsabilités) ; ces dimensions se retrouvent en partie dans les écrits scientifiques portant sur les effets fonctionnels reliés au TPL.

Par la suite, des groupes de discussion focalisée ont été réalisés auprès de personnes ayant un TPL. Ceux-ci avaient pour but de comprendre les liens existant entre le fait d'avoir un TPL et les difficultés vécues dans le fonctionnement quotidien, un sujet sur lequel peu de données sont actuellement disponibles dans les écrits scientifiques. Les sujets traités dans les groupes de discussion étaient les dimensions et sous-dimensions du fonctionnement choisis par les professionnels de la santé lors de la consultation préalable, comme points de départ. Quelques exemples des dimensions et sous-dimensions traitées apparaissent dans le tableau 12.2. Les participants (les personnes avec un TPL) ont été invités à se prononcer sur les différentes sphères du fonctionnement présentées afin de déterminer si oui ou non cette dimension était affectée dans leur quotidien. De plus, les participants devaient aussi expliquer de quelle manière le TPL affecte leur fonctionnement ainsi que les stratégies développées afin de s'adapter à ces difficultés.

Tableau 12.2.
Liste des sujets de discussion

Dimensions	Sous-dimensions
1. Fonctions du tempérament	Extraversion, amabilité, stabilité psychique, ouverture aux expériences, optimisme
2. Fonctions de l'énergie et des pulsions	Motivation, dépendances, contrôle des pulsions
3. Fonctions émotionnelles	Pertinence des émotions, maîtrise des émotions
4. Prise de décision	*Aucune sous-dimension relevée par les répondants*
5. Réalisation de la routine quotidienne	Gérer la routine quotidienne, mener à bien les routines, gérer son propre niveau d'activité

2.3. Recrutement

Les participants ont été recrutés par les médecins responsables de la clinique des troubles relationnels et de la personnalité de l'Institut en santé mentale de Montréal, qui avaient pour tâche d'identifier des personnes présentant un diagnostic unique de TPL, ainsi que de cibler des personnes présentant des profils variés de la pathologie et des effets dans le fonctionnement de la vie quotidienne. Une attention particulière a été accordée à la sélection

des participants pour représenter un vaste éventail d'expériences face à la maladie. La richesse apportée par de multiples expériences permettait une compréhension plus large et plus en profondeur du concept étudié. À partir de ces critères, douze personnes ont été initialement recrutées et ensuite contactées par la responsable de la recherche. À la suite des explications de la nature de la participation au projet de recherche, huit personnes ont accepté de participer à différents groupes de discussion. Les participants ont reçu une compensation financière.

2.4. Ressources

Un repas léger a aussi été servi durant le groupe puisque celui-ci devait se tenir de 17 h à 19 h compte tenu des disponibilités des personnes. Les rencontres se sont déroulées à l'intérieur des locaux de l'établissement hospitalier dans une salle ventilée, où une grande table carrée et des chaises confortables étaient disponibles. Un enregistreur numérique a aussi été nécessaire afin d'enregistrer l'intégralité des rencontres en format audio. Quant aux ressources humaines nécessaires, la responsable du projet de recherche a animé les rencontres et la coanimation a été confiée à un agent de planification de la recherche de l'établissement, tous deux ayant de l'expérience dans l'animation de groupes.

2.5. Déroulement

Lors de la présentation initiale du déroulement des séances, il a été expliqué aux huit participants que le nombre total de rencontres ne pouvait être déterminé à l'avance, mais que plusieurs rencontres seraient nécessaires afin de couvrir l'ensemble des sujets de discussion. Chaque séance a permis de traiter de plusieurs sujets, chaque sujet étant exploré jusqu'à l'épuisement d'éléments nouveaux à la discussion. Les participants décidaient d'eux-mêmes si le sujet s'était tari ou non. Certains sujets ont été repris lors de rencontres subséquentes en raison de l'intérêt de certaines personnes d'ajouter de nouveaux éléments apparus à la suite de leur réflexion personnelle. En raison des disponibilités variables des participants, plusieurs groupes différents ont été formés (p. ex., groupe A = participants 1-2-3-4; groupe B = 5-6-7-8; groupe C = 1-2-7-8, etc.). La composition de chacun des groupes a donc été déterminée en fonction des disponibilités du plus grand nombre, la taille variant lors de chaque rencontre entre quatre et huit participants.

Le rôle de l'animateur était de présenter les règles de fonctionnement du groupe, tels la confidentialité, le droit de parole, l'importance de respecter les points de vue divergents, etc. L'animateur devait aussi s'assurer que tous puissent s'exprimer librement sur chacune des dimensions abordées. Ainsi, certaines interventions ont été nécessaires afin d'encourager les participants à s'exprimer sur un sujet ; ceux-ci ont admis avoir apprécié être interpellés et que l'on souligne l'importance de l'apport de leurs points de vue, même si différents. De son côté, le coanimateur s'est assuré que le matériel nécessaire à l'enregistrement soit fonctionnel et que les ressources matérielles favorisant une bonne ambiance (boissons, nourriture, aération de la salle, etc.) soient mises en place. Le coanimateur avait aussi pour rôle de contribuer à l'établissement d'une ambiance propice à la discussion de groupe ; à cet effet, il est intervenu quelques fois afin de tempérer des échanges plus houleux et de valider des opinions divergentes, mais complémentaires.

Nous avons rencontré un problème lors de l'animation des groupes de discussion focalisée, lié à la capacité d'une participante à respecter le cadre établi. En effet, malgré le processus de sélection lors de la référence par le médecin traitant, il s'est avéré que cette personne traversait une période de plus grande vulnérabilité au moment de sa participation à un groupe de discussion. Lors de sa participation au groupe (une seule séance), des difficultés à se relier aux autres participants et à tolérer les opinions différentes des siennes ont été observées, ce qui a nui à l'établissement d'une ambiance propice à la discussion. Cette personne a par conséquent été invitée à se retirer du processus de consultation pour partager ses idées lors de rencontres individuelles. Nous avons fait l'hypothèse que ses difficultés à interagir avec les autres participants étaient peut-être en lien avec la nature même du trouble de la personnalité limite, et qu'il était donc nécessaire d'enrichir notre compréhension des effets fonctionnels et sociaux du TPL. Il a donc été convenu avec la participante qu'elle contacterait elle-même les chercheurs au moment jugé opportun afin d'offrir sa contribution sur une base individuelle et donner son opinion sur les différentes dimensions proposées.

2.6. Collecte des données

Afin de recueillir les données, un guide d'entrevue a été conçu par l'équipe de recherche. Le canevas apparaît dans l'encadré 12.1.

Encadré 12.1.
Extraits du guide d'entrevue

RÔLE DE L'ANIMATEUR – MOT D'INTRODUCTION	
	Bonjour à tous, *D'abord, je vous remercie de votre intérêt pour ce projet. Le projet a pour but de cibler les dimensions de votre fonctionnement qui peuvent être difficiles pour vous ou représenter un défi au quotidien. Dans le cadre de cet entretien de groupe, nous voulons essentiellement connaître votre expérience de certains aspects du fonctionnement au quotidien. Les échanges vont durer un peu moins de deux heures. Il n'y a pas de bonnes ou de mauvaises réponses, uniquement des points de vue différents. Il n'est pas nécessaire non plus de parvenir à un consensus. Tous doivent se sentir à l'aise d'émettre leur opinion.* *Pour faciliter la discussion, je vous demande de parler assez fort et une seule personne à la fois. Nous enregistrons parce que nous ne voulons perdre aucune des opinions émises. Aux fins de la discussion, nous n'utiliserons que vos prénoms, mais ceux-ci seront remplacés par des noms fictifs lorsque nous rédigerons le compte rendu, soyez assurés que votre anonymat est préservé.* *Si vous en ressentez le besoin (parce qu'un sujet abordé a réveillé des émotions vives, par exemple), il est bien entendu possible de venir m'en parler dès la fin du groupe ou de prendre contact avec moi plus tard ; nous pourrons ainsi nous assurer que vous aurez accès à du soutien, en vous dirigeant vers un organisme ou en vous encourageant à en discuter avec votre équipe traitante, selon votre préférence.*
QUESTION 1 *10 minutes*	La première question requiert un tour de table. C'est une question « brise-glace ». *Croyez-vous qu'avoir un trouble de la personnalité limite influence votre façon de fonctionner dans votre vie de tous les jours ?*
QUESTION 2 ET SUIVANTES *90 minutes*	Les questions suivantes ont toutes la même forme. Elles visent à couvrir chacune des dimensions du fonctionnement telles qu'elles apparaissent dans le tableau 12.3, et à établir avec les participants si cette dimension est atteinte et de quelle façon elle l'est. *Diriez-vous que cet aspect (dimension x du fonctionnement : p. ex., s'organiser au quotidien) représente une difficulté pour vous ?* *Pouvez-vous décrire de quelle façon cette dimension a pu être affectée lors de vos pires moments ? Et quels moyens avez-vous développés pour faire face à cette difficulté ?*
CONCLUSION *5 à 10 minutes*	Le résumé de la discussion peut être fait par l'animateur.
RÉSUMÉ DES ÉCHANGES *5 à 10 minutes*	Après avoir présenté le résumé, poser les questions suivantes : *Est-ce que quelque chose a été oublié ?* *Est-ce que quelqu'un parmi vous souhaite ajouter autre chose, apporter une précision ?*

2.7. Analyse et résultats

Au total, neuf rencontres de discussion de groupe ont été nécessaires afin d'obtenir la saturation des données. À la neuvième rencontre, les participants eux-mêmes ont dit avoir l'impression de se répéter et que le contenu des discussions revenait au même d'une fois à l'autre. Les huit participants ont été conviés aux neuf rencontres, mais tous n'ont pu offrir le même niveau d'engagement dans le processus ; dans les faits, les neuf rencontres ont donné lieu à des compositions différentes d'une fois à l'autre, variant de quatre à huit participants (comme décrit précédemment). Bien que la théorie indique qu'il est préférable de soumettre chaque sujet de discussion à un minimum de trois groupes différents, il n'a pas été possible de le faire ici compte tenu de l'accès restreint à un petit nombre de participants et de ressources financières limitées. Cependant, l'atteinte de la saturation des données, obtenue auprès de personnes présentant un éventail très large d'expériences de la maladie, rassure sur la fiabilité des données issues du processus de collecte des données. Les neuf rencontres ont été enregistrées, transcrites, lues à plusieurs reprises et codées à l'aide d'une thématisation en continu. Cette démarche a été adoptée, car elle permet une analyse fine et riche du corpus (Paillé, 2006). Elle consiste en une démarche ininterrompue d'attribution de thèmes et, simultanément, de construction d'un arbre thématique (Paillé, 2006). Afin d'assurer la crédibilité de la thématisation, un codage indépendant avec une fidélité interjuges de plus de 90 % entre l'animateur et le coanimateur a été réalisé. Les résultats des analyses ont également été soumis aux huit participants du groupe de discussion, lors d'une rencontre de groupe subséquente au processus d'analyse afin d'obtenir leurs rétroactions (Côté et Turgeon, 2002).

À la fin de chaque rencontre de groupe, une discussion informelle entre l'animateur et le coanimateur a permis de dégager les grandes lignes de l'analyse thématique. L'une des premières constatations ayant émergé de cette période de discussion est que les participants présentaient des expériences très variées de la maladie, ceci venant confirmer une certaine représentativité de l'échantillon choisi. Malgré cela, les participants ont pu trouver des similitudes et des significations communes afin d'expliquer ces divergences et ces écarts dans les expériences vécues. Une autre constatation faite par les animateurs a été que les participants ont réussi par eux-mêmes à faire ressortir toute la gradation possible du fonctionnement, allant du pire au meilleur, et ce, à partir de leurs expériences passées. En effet, les participants faisaient des descriptions très claires des moments de leur vie où leur fonctionnement était détérioré ou désorganisé, et étaient aussi en mesure de décrire quelles stratégies ils avaient mises en œuvre afin de contourner ces difficultés ou d'y faire face et ainsi atteindre un meilleur fonctionnement.

Lors de l'analyse des données, plusieurs thèmes ont été retenus au fur et à mesure par le chercheur principal (analyse séquentielle), et ce, pour chacune des dimensions présentées. Dans l'ensemble, les thèmes étaient explicatifs des difficultés éprouvées au quotidien par les personnes ayant un TPL. En effet, les participants ont voulu donner un sens et des explications à leurs difficultés au quotidien plutôt que de se limiter à simplement confirmer ou infirmer leur difficulté par rapport à la dimension présentée. Pour chacune de ces dimensions, des explications ont été données par les participants, et plusieurs thèmes ont émergé. Par exemple, nous présentons ici un thème relatif aux stratégies pour faire face aux difficultés associées à la dimension de gestion du temps, soit « se donner une marge de manœuvre ». Cette stratégie semblait être une façon de ne pas vivre des échecs au quotidien pouvant miner considérablement l'estime de soi. Les personnes ayant un TPL ont souvent des exigences personnelles élevées et ont dit préférer ne pas s'engager dans certaines activités ou envers des personnes afin d'éviter toute possibilité d'un échec ou de décevoir l'entourage. Une fois libérée de la pression de réussir à la perfection une activité amorcée, la personne sent qu'il est alors possible d'accomplir certaines activités de façon plus dégagée. Voici un exemple associé à cette stratégie : « *Je garde mon énergie pour le travail et le reste, il faut que ce soit pas mal de la marge de manœuvre, que je puisse aménager mon temps. Si je donne un rendez-vous à une amie, je sais que je peux l'annuler. Ce n'est pas si grave que ça parce que ce n'est pas le travail. Je priorise les choses les plus importantes.* »

En plus de thèmes associés à chacune des dimensions et sous-dimensions, quatre thèmes généraux, communs à l'ensemble des dimensions, ont été dégagés de l'analyse de contenu des groupes de discussions focalisée. Il s'agit de : « la stratégie efficace pour faire face aux difficultés », « la stratégie nuisible pour faire face aux difficultés », « les pires impacts de mon TPL sur le fonctionnement », et « fonctionner au mieux avec le TPL ».

Par ailleurs, les interactions entre les participants ont aussi été observées, notées et ont fait l'objet d'analyse en continu. Le tableau 12.3 présente un bref extrait d'échanges entre les participants à propos de la dimension « gérer la routine quotidienne » ; la colonne de gauche indique les thèmes d'analyse du contenu et la colonne de droite, les thèmes d'analyse des interactions.

Tableau 12.3.
Thèmes d'analyse du contenu et des interactions

Analyse du contenu	Éléments des échanges, tirés des comptes rendus intégraux (*verbatim*)	Analyse des interactions
	Animateur : *C'est lorsqu'il faut changer la routine que c'est difficile ?*	
Comportement nuisible au fonctionnement.	**P-1 :** *C'est surtout que je n'avais pas de routine avant, parce que je ne faisais rien.*	
Stratégie : limiter les activités productives.	*J'ai été un an sans travailler, sans aller à l'école, à seulement faire mes thérapies.*	
Amélioration du fonctionnement.	*Ça a vraiment été quelque chose de reprendre la vie sociale, de réétablir une routine. Ça a pris un bon six mois.*	
	P-2 : *Ce qu'elle dit regroupe plusieurs choses dont on a parlé j'ai l'impression.*	Réalité commune
	P-3 : *Je ne sais pas si ça a rapport… mais…*	Distanciation
Stratégie : limiter le temps consacré au travail (temps partiel).	*Je ne suis pas capable d'avoir une routine comme bien du monde, de gérer la routine de 40 heures en plus des enfants comme les autres font.*	Distanciation
Stratégie : limiter le temps consacré au travail (temps partiel).	*Je suis incapable, je trouve ça trop. Je n'ai pas d'enfants, je fais trente heures puis c'est bien assez.*	
	Je ne sais pas comment les autres font.	Réalité différente
	P-4 : *Moi, je vis totalement le contraire.*	Réalité différente
Être surengagé.	*Je travaille depuis janvier sept jours par semaine. J'ai un adolescent à la maison. J'ai mes animaux à m'occuper. Je fais toutes mes tâches toute seule dans la maison, et je n'ai pas d'aide.*	Réalité différente
Stratégie d'isolement.	Mais… je n'ai pas de vie sociale. J'ai coupé la vie sociale parce que je suis sauvage de toute façon.	
Ne pas demander d'aide entraîne insatisfaction et expression inadéquate de la colère.	*Mais y'a des fois où je ne suis pas contente, je suis marabout* [irritable]. *Je vais crier après mon copain, mais il me dit : « Arrête d'en faire trop ! » Quand je suis moins en forme, et que je trouve que j'en ai trop à faire, je vais engueuler les autres qui sont dans la maison parce qu'ils sont assis et ils prennent leur temps.*	

(*suite*)

Tableau 12.3.
Thèmes d'analyse du contenu et des interactions (*suite*)

Analyse du contenu	**Éléments des échanges, tirés des comptes rendus intégraux (*verbatim*)**	**Analyse des interactions**
	P-5 : *Ah ça, ça me ressemble ! ! !* [rires]	Réalité commune
Ne pas demander d'aide entraîne insatisfaction.	*Gérer les enfants, la routine, quand je suis fatiguée, c'est de plus en plus difficile. Partager les tâches… j'ai de la misère à le dire. Je vais le dire, mais pas de la bonne façon.*	
Insatisfaction entraîne expression inadéquate de la colère.	**P-4 :** *Moi, je crie fort*	Réalité commune
Expression inadéquate du besoin réel.	**P-5 :** *On va le dire, mais pas de la façon que l'autre va comprendre notre détresse.*	Réalité commune
Stratégie afin d'éviter la détresse émotionnelle.	**P-4 :** *On dirait que la journée que je vais m'arrêter, je vais m'écraser. Alors, je ne veux pas arrêter*.	Réalité différente

L'analyse de l'ensemble des rencontres a permis de dégager quatre types d'interactions : 1) les manifestations de reconnaissance d'une réalité commune ; 2) les manifestations de reconnaissance d'une réalité différente ; 3) les tentatives de normalisation ; et 4) la distanciation. Les manifestations d'une réalité commune ont pris la forme de signes de reconnaissance tels qu'une écoute accrue, des signes non verbaux d'approbation, des indices verbaux (p. ex., «*Moi, c'est exactement pareil*»), ainsi que des éléments verbaux afin d'ajouter un élément ou le préciser (p. ex., la réponse d'un autre participant à cette dernière intervention : «*Avec ce que tu me dis, je vois qu'on vit la même chose, mais moi quand tu parles, je me reconnais il y a vingt ans… j'ai changé, j'ai évolué, mais le fond est pareil.*» Pour toutes les dimensions jugées atteintes par les participants, ces manifestations ont pu être relevées et étaient fréquentes.

Les manifestations d'une réalité différente ont pris la forme de signes exprimant l'intérêt tels que des demandes de clarification, des signes verbaux d'étonnement (p. ex., «*Ah oui ?*»), des tentatives de négation des différences (p. ex., «*On dit la même chose au fond…*»), ainsi que des tentatives d'explications de ces différences malgré le diagnostic commun (p. ex., «*Je me rends compte qu'on est dans les extrêmes quand on est TPL, c'est tout un ou tout l'autre…*»). Les sujets de discussion où ces manifestations sont apparues ont mené à des échanges riches sur la pathologie et sur les facteurs explicatifs de réalités différentes.

Les tentatives de normalisation ont pris la forme de questionnements concernant la pertinence d'une dimension choisie. Par exemple, les participants se sont demandé si ce n'était pas seulement normal d'avoir des difficultés à gérer une crise et si toutes les personnes, avec ou sans un TPL, n'éprouvaient pas de difficultés par rapport à cet aspect. Les sujets ou les dimensions pour lesquels il y avait ces questionnements ont donné lieu à de longs échanges, et la plupart de ces dimensions ont finalement été jugées significativement atteintes selon leurs points de vue.

Enfin, les manifestations de distanciation ont pris la forme de retrait, de signes non verbaux tels le silence, le regard ailleurs, des hochements de tête, ainsi que l'expression verbale de distance face au groupe (p. ex., «*Je ne suis pas certaine que c'est en lien ce que je vais dire, mais j'ai l'impression d'être la seule différente ici.*»). Ces manifestations ont donné lieu à des efforts accrus de la part des autres participants afin de donner un sens à la différence et ont permis d'élaborer une compréhension de la maladie.

Dans cette étape de l'étude, les groupes de discussion focalisée ont donné lieu à des échanges riches et en profondeur sur les éléments pouvant permettre une compréhension du fonctionnement au quotidien des personnes présentant un TPL. L'analyse des interactions ayant eu lieu entre les participants a permis de mettre en perspective les réponses données individuellement et de mener à une compréhension plus complète des liens existant entre le TPL et ses effets dans la vie de tous les jours. Les thèmes principaux issus de l'analyse du contenu de ces échanges de groupe serviront de base conceptuelle à l'élaboration d'un outil de mesure du fonctionnement pour ces personnes.

CONCLUSION

Ce chapitre a permis un survol des différents aspects théoriques et pratiques relatifs à l'élaboration et à la conduite des groupes de discussion focalisée. L'aspect central de cette méthode de recherche est d'apprendre des participants, de leurs expériences et opinions. À cette fin, l'ambiance se doit d'être conviviale et la composition du groupe, assez homogène pour permettre le confort et favoriser les échanges. Toutefois, un certain degré d'hétérogénéité (p. ex., des gens d'âges et de sexes différents) doit être recherché chez les participants, afin d'assurer la validité des données. Le nombre de séances et de participants ne peut être défini à l'avance: il demande à être raisonné et justifié pour chaque projet particulier, en fonction des buts et des objectifs, de même que des contraintes et du temps nécessaire pour atteindre la saturation des données. Une attention particulière doit être apportée à la phase préparatoire, qui revêt une grande importance afin d'assurer le

bon déroulement des groupes ainsi que la fiabilité des données obtenues. Selon Morgan (1998), un spécialiste de la méthode, il n'existe pas de meilleure façon d'apprendre à faire des groupes de discussion focalisée qu'en les expérimentant.

RÉFÉRENCES

ANDERSEN, A. (1991). *A Process Guidebook: Tools and Techniques for Effective Facilitation*, St. Charles, Arthur Andersen & Co.

BARIBEAU, C. (2009). « Analyse des données des entretiens de groupe », *Recherches qualitatives*, vol. 28, nº 1, p. 133-148.

BARIBEAU, C. et M. GERMAIN (2010). « L'entretien de groupe : considérations théoriques et méthodologiques », *Recherches qualitatives*, vol. 29, nº 1, p. 28-49.

BARIBEAU, C., J. LUCKERHOFF et F. GUILLEMETTE (2010). « Introduction : les entretiens de groupe », *Recherches qualitatives*, vol. 29, nº 1, p. 1-4.

CÔTÉ, L. et J. TURGEON (2002). « Comment lire de façon critique les articles de recherche qualitative en médecine », *Pédagogie médicale*, vol. 3, nº 2, p. 81-90.

DALKEY, N. et O. HELMER (1963). « An experimental application of the Delphi method to the use of expert », *Management Science*, vol. 9, nº 3, p. 458-467.

DAWSON, S. et L. MANDERSON (1993). *Le manuel des groupes focaux, méthodes de recherche en sciences sociales sur les maladies tropicales*, New York, PNUD/Banque mondiale/OMS.

GEOFFRION, P. (2009). « Le groupe de discussion », dans B. Gauthier (dir.), *Recherche sociale. De la problématique à la collecte des données*, 5e éd., Québec, Presses de l'Université du Québec, p. 391-414.

GRONKJAER, M., T. CURTIS, C. DE CRESPIGNY et C. DELMAR (2011). « Analysing group interaction in focus group research : Impact on content and the role of the moderator », *Qualitative studies*, vol. 2, nº 1, p. 16-30.

HOLLANDER, J.A. (2004). « The social contexts of focus groups », *Journal of Contemporary Ethnography*, vol. 33, nº 5, p. 602-637.

JOVCHELOVITCH, S. (2004). « Contextualiser les focus groups : comprendre les groups et les cultures dans la recherche sur les représentations », *Bulletin de psychologie*, vol. 57, nº 3, p. 245-261.

KITZINGER, J. (1995). « Introducing focus groups », *British Medical Journal*, vol. 311, p. 299-302.

KITZINGER, J., I. MARKOVA et N. KALAMPALIKIS (2004). « Qu'est-ce que les focus groups ? », *Bulletin de psychologie*, vol. 57, nº 3, p. 237-243.

KRUEGER, R.A. (1994). *Focus Groups. A Practical Guide for Applied Research*, 2e éd., Thousand Oaks, Sage Publications.

KRUEGER, R.A. (1998). « Analyzing and reporting focus group results », *Focus Group Kit 6*, Thousand Oaks, Sage Publications, p. 36, 46, 69.

KRUEGER, R.A. et M.A. CASEY (2009). *Focus Groups: A Practical Guide for Applied Research*, Thousand Oaks, Sage Publications.

LASCH, K.E., P. MARQUIS, M. VIGNEUX, L. ABETZ, B. ARNOULD, M. BAYLISS, B. CRAWFORD et K. ROSA (2010). « PRO development: Rigorous qualitative research as the crucial foundation », *Quality of Life Research*, vol. 19, p. 1087-1096.

MORGAN, D.L. (1998b). « The focus group guidebook », *Focus Group Kit 1*, Thousand Oaks, Sage Publications, p. 30.

MORGAN, D.L. (2004). « Focus groups », dans S.N. Hesse-Biber et P. Leavy (dir.), *Approaches to Qualitative Research*, New York, Oxford University Press, p. 263-285.

ORGANISATION MONDIALE DE LA SANTÉ – OMS (2001). *Classification internationale du fonctionnement, du handicap et de la santé (CIF)*, Genève, OMS.

PAILLÉ, P. (2006). « Qui suis-je pour interpréter ? », dans P. Paillé (dir.), *La méthodologie qualitative, postures de recherche et travail de terrain*, Paris, Armand Colin, p. 99-123.

POWELL, R.A. et H.M. SINGLE (1996). « Focus groups », *International Journal for Quality in Health Care*, vol. 8, n° 5, p. 499-504.

PUCHTA, C. et J. POTTER (2004). *Focus Group Practice*, Londres, Sage Publications.

RABIEE, F. (2004). « Focus-group interview and data analysis », *Proceedings of the Nutrition Society*, vol. 63, p. 655-660.

RICHARDSON, C.A. et F. RABIEE (2001). « A question of access: An exploration of the factors influencing the health of young males aged 15–19 living in Corby and their use of health care services », *Health Education Journal*, vol. 60, n° 2, p. 3-6.

SMITHSON, J. (2000). « Using and analysing focus groups: Limitations and possibilities », *International Journal of Social Research Methodology*, vol. 3, n° 2, p. 103-119.

TOURÉ, E.H. (2010). « Réflexion épistémologique sur l'usage des *focus groups :* fondements scientifiques et problèmes de scientificité », *Recherches qualitatives*, vol. 29, n° 1, p. 5-27.

VICSEK, L. (2007). « A scheme for analyzing the results of focus groups », *International Journal of Qualitative Methods*, vol. 6, n° 4, p. 20-34.

CHAPITRE 13

L'APPROCHE DELPHI
Application dans la conception d'un outil clinique en réadaptation au travail en santé mentale

Valérie Tremblay-Boudreault
Clermont E. Dionne

FORCES

- Elle permet de guider un groupe d'experts vers un consensus et d'arriver à une recommandation finale, tout en favorisant l'acceptabilité des résultats par les participants.
- Elle est efficace et peu coûteuse comparée à plusieurs autres.
- L'anonymat favorise l'expression et la considération des opinions de tous les participants.

LIMITES

- Il n'y a pas de critère pour sélectionner les experts, pour déterminer la représentativité de l'échantillon ni pour établir sa taille.
- Il n'y a pas de consensus sur ce qu'est un consensus.
- Les pertes au suivi sont souvent nombreuses s'il y a plusieurs tours de consultation.

1. PRÉSENTATION DE LA MÉTHODE DELPHI

Certaines questions de recherche n'ont qu'une seule réponse et les chercheurs la trouvent avec les bons outils, à coup d'efforts ciblés. D'autres questions ont plusieurs réponses possibles, car les évidences sont insuffisantes ou contradictoires, les variables d'intérêt sont intangibles ou vagues, ou les méthodes de recherche quantitatives ou qualitatives traditionnelles ne parviennent pas à fournir une solution qui soit supérieure d'une quelconque façon (Duffield, 1988 ; Hasson, Keeney et McKenna, 2000 ; Jones et Hunter, 1995 ; Steurer, 2011). Dans de telles situations, on peut réduire le nombre de réponses en demandant à des « experts » de la question à l'étude de travailler pour atteindre un consensus. La réponse finale issue de ce processus sera supérieure aux autres réponses possibles sur au moins deux aspects : 1) elle aura été préférée conjointement par plusieurs experts, ce qui en augmentera l'acceptabilité, et 2) elle offrira les avantages liés à la standardisation (meilleure validité externe, notamment). Plusieurs méthodes permettent d'atteindre un consensus d'experts (McMurray, 1994), incluant les rencontres de comités, la technique du groupe nominal (Van de Ven et Delbecq, 1972), les entrevues en face à face, les groupes de discussion focalisée (*focus groups*) (voir le chapitre 12 de cet ouvrage), les rencontres de remue-méninges (*brainstorming*) (Couper, 1984) et la RAND/UCLA Appropriateness Method (une forme de rencontre hautement structurée) (Fitch *et al.*, 2001). Cependant, ces méthodes sont souvent coûteuses, elles ont tendance à donner plus de poids aux arguments de certains experts (le facteur de ralliement [*bandwagon effect*]) et ne permettent pas la plupart du temps de mesurer adéquatement les résistances et les désaccords des participants (Beech, 1999 ; Steurer, 2011).

La méthode Delphi (le terme fait référence à l'oracle de la ville antique grecque de Delphi [Broad, 2006 ; Steurer, 2011]) a été développée par Dalkey et ses collaborateurs à la Rand Corporation à la demande de l'armée de l'air américaine au début des années 1950, pour obtenir des consensus d'experts concernant les conséquences envisagées de la guerre froide (Couper, 1984 ; Dalkey et Helmer, 1963 ; Linstone et Turoff, 1975). Elle est fondée sur le principe que « plusieurs têtes valent mieux qu'une » (Couper, 1984). En sciences de la santé, cette méthode est utilisée depuis le milieu des années 1970, particulièrement en sciences infirmières (Powell, 2003). Elle peut être définie comme « une méthode efficiente de structuration du processus de communication, permettant à un groupe d'individus de gérer ensemble un problème complexe » (Linstone et Turoff, 1975, p. 3 ; traduction libre). Les fondements absolus de cette méthode, qui la

caractérisent, sont : 1) plusieurs itérations de collecte d'information, avec un *certain*[1] retour d'information aux participants, 2) une *certaine* évaluation du consensus, et 3) un *certain* niveau d'anonymat pour les réponses individuelles (Couper, 1984 ; Linstone et Turoff, 1975).

Vue comme une méthode de structuration du processus de communication, la méthode Delphi peut être appliquée à tous les domaines de recherche. Les circonstances particulières qui justifient l'utilisation de cette méthode comprennent : 1) un problème qui ne peut pas être résolu par un devis de recherche particulier et qui peut bénéficier du jugement subjectif sur une base collective ; 2) plusieurs individus qui doivent interagir pour résoudre le problème ; 3) ces individus ne se connaissent pas ou, s'ils se connaissent, n'ont pas d'histoire de communication antérieure à ce sujet ; 4) les contraintes de temps et de coûts rendent les rencontres fréquentes irréalisables ; 5) l'efficience de rencontres en face à face n'est pas optimale ; 6) les divergences d'opinions entre les individus ou la nature du sujet sont telles que l'anonymat est requis ; 7) l'hétérogénéité des participants doit être préservée, tout en évitant la domination par un groupe ou par des personnalité fortes (Couper, 1984 ; Jones et Hunter, 1995 ; Linstone et Turoff, 1975 ; Powell, 2003).

Dans les sections qui suivent, nous présenterons la procédure de collecte, d'analyse et d'interprétation des résultats, les limites de la méthode Delphi, les logiciels et autre matériel nécessaires pour utiliser cette méthode, ainsi qu'un exemple d'application de la méthode Delphi en santé mentale au travail.

1.1. Procédure de collecte, d'analyse et d'interprétation des résultats

Le processus Delphi *classique* (qui fait l'objet de diverses modifications dans la plupart des études) se déroule généralement en cinq phases, présentées ici.

1.1.1. Mise sur pied d'un panel d'individus informés (« experts »)

La sélection (Hasson *et al.*, 2000) des membres de ce panel, à qui on demandera individuellement et de façon anonyme de faire des suggestions sur le sujet d'intérêt, est cruciale (Campbell *et al.*, 1999 ; Duffield, 1993 ; Linstone et Turoff, 1975 ; Steurer, 2011). Une approche fréquemment utilisée consiste à sélectionner un ou deux experts et à leur demander d'en proposer d'autres à leur tour (méthode « boule de neige ») (Steurer, 2011). Cette façon de

1. Le terme *certain* est utilisé spécifiquement par Linstone et Turoff. La définition de *certain* demeure large encore aujourd'hui.

faire, quoique pratique, peut amener les experts à choisir des collègues qui ont les mêmes opinions qu'eux. D'autres méthodes se veulent plus systématiques, par exemple la sélection de tous les premiers auteurs d'articles sur un sujet particulier au cours d'une période donnée (Steurer, 2011). Les écrits scientifiques indiquent que les groupes hétérogènes sont préférables aux groupes homogènes (Bantel, 1999 ; Couper, 1984 ; Powell, 2003). Les membres du panel d'experts seront choisis selon leur expertise et leur désir d'y participer (Goodman, 1987 ; Hasson *et al.*, 2000). Le nombre de participants peut varier de dix à plusieurs centaines, voire des milliers d'individus (Powell, 2003). Certains chercheurs considèrent que la fidélité des résultats augmente avec le nombre (Couper, 1984) ou le niveau de connaissance du sujet par les participants (Akins, Tolson et Cole, 2005), alors que d'autres pensent que les écrits scientifiques concernant l'effet du nombre et du type d'experts sur la fidélité et la validité des résultats sont minces (Mullen, 2003 ; Powell, 2003).

Bien que les difficultés liées à la sélection des experts constituent une faiblesse de la méthode Delphi, elles ne lui sont pas exclusives. Par exemple, la technique du groupe nominal partage la même faiblesse (Jones et Hunter, 1995 ; Keeney, Hasson et McKenna, 2006 ; Linstone et Turoff, 1975).

1.1.2. Exploration du sujet et retour des questionnaires

Lors de cette phase, chaque participant répondra à des questions de base sur le sujet d'intérêt et pourra ajouter de l'information qu'il juge pertinente (p. ex., « Y a-t-il d'autres éléments de cette problématique que nous n'avons pas considérés ? »). Le rôle du premier questionnaire sera souvent de déterminer les éléments qui feront l'objet des tours de consultation suivants. Il comportera souvent des questions ouvertes qui contribueront à enrichir les résultats (Powell, 2003). Cependant, si le premier questionnaire est trop général et peu structuré, les participants peuvent avoir de la difficulté à répondre ou encore, les réponses peuvent ne pas être pertinentes aux besoins de l'étude. Il peut donc être plus judicieux de commencer le premier questionnaire avec des questions précises (Duffield, 1988). Un maximum de 25 questions est recommandé (Couper, 1984). Il est possible d'ajouter des questions, mais le succès dépendra alors de leur organisation, de leur logique, de la facilité à répondre et de l'espace disponible pour les commentaires (Duffield, 1988 ; Linstone et Turoff, 1975). Il est important que les questions posées et les instructions soient très claires afin que chacun les interprète de la même façon (Linstone et Turoff, 1975). Une étude pilote préalable sera très utile à cet effet (Clibbens, Walters et Baird, 2012 ; Duffield, 1988 ; Hasson *et al.*, 2000 ; Powell, 2003). Il faut savoir que l'opinion des

experts sera probablement plus influencée par des événements récents que par le passé lointain ou les perspectives futures (effet *discounting the future*) (Couper, 1984; Linstone et Turoff, 1975).

À l'origine, les questionnaires étaient envoyés par la poste. De nos jours, il est de plus en plus fréquent d'utiliser des questionnaires Web pour les avantages qu'ils offrent en matière d'économie de temps et de ressources, et de flexibilité (Cole, Donohoe et Stellefson, 2013; Deshpande et Shiffman, 2003; Linstone et Turoff, 1975; Steurer, 2011). Ils donnent toutefois lieu à des pertes au suivi plus nombreuses que les questionnaires envoyés par la poste, notamment en raison des difficultés techniques, qui semblent incontournables (Cole *et al.*, 2013). Il peut être judicieux d'utiliser les deux approches conjointement, par exemple des questionnaires Web et des rappels par la poste (Cole *et al.*, 2013).

1.1.3. Synthèse des positions individuelles et définition des zones d'accord et de désaccord

À cette phase repose l'élément méthodologique le plus sensible de la méthode Delphi: la définition du consensus, c'est-à-dire l'accord entre les participants (p. ex., une mesure de tendance centrale d'une réponse quantitative des participants – comme la moyenne arithmétique – ou encore la stabilité des opinions des participants d'un tour de consultation à l'autre, ou enfin, une mesure subjective de l'opinion générale) (Holey *et al.*, 2007). Les chercheurs doivent définir cet accord de façon opérationnelle *avant* de commencer la collecte d'information (Keeney *et al.*, 2006). Si un indicateur quantitatif est utilisé, le seuil minimal définissant l'accord doit être déterminé à l'avance et la même valeur doit être utilisée lors de chaque tour de consultation. La définition de l'accord aura une incidence particulière sur le nombre de tours de consultation requis pour atteindre un consensus.

Paradoxalement, il n'existe pas de consensus sur la définition d'un consensus. Les chercheurs doivent donc décider *a priori* d'un critère (Hasson *et al.*, 2000; Keeney *et al.*, 2006; Linstone et Turoff, 1975). Certains chercheurs considèrent qu'un accord de 51% est suffisant, alors que d'autres ont proposé comme critère 70%, 80% ou même 100%, d'autres encore allant jusqu'à remettre en question la valeur d'un critère quantitatif, proposant plutôt de continuer le processus de consultation jusqu'à la stabilité des réponses (Hasson *et al.*, 2000; Keeney *et al.*, 2006; Mullen, 2003; Powell, 2003). Holey *et al.* (2007) ont démontré que la comparaison de certaines statistiques quantitatives (médiane, moyenne, étendue, écart-type, coefficients Kappa, nombre de commentaires) entre les tours de consultation témoigne bien de la convergence des positions vers un consensus. D'autres

définitions opérationnelles peuvent aussi être utilisées, incluant celle recommandée dans le cadre de la RAND/UCLA Appropriateness Method (RAM) (Fitch *et al.*, 2001), dont une application sera décrite dans la seconde partie du présent chapitre.

1.1.4. Exploration des désaccords et retour aux participants

Le second tour de consultation et les suivants seront généralement plus précis que le premier. Ils visent à quantifier les résultats obtenus lors du tour précédent et à retourner cette information aux participants (Powell, 2003). Il est d'usage de fournir de l'information à chaque participant sur sa position et celle du groupe pour chacune des questions. Chaque répétition de cette phase constitue un « tour de consultation » (*round*). L'itération et le retour d'information permettront aux participants de reconsidérer leur position à la lumière des autres opinions, ce qui est facilité par l'anonymat (Jones et Hunter, 1995), de tenir compte d'informations supplémentaires ou de nouveaux arguments, et de converger collectivement vers un consensus (Steurer, 2011). Les statistiques transmises à chacun des participants pourront prendre différentes formes. Elle seront cependant présentées de la même façon d'un tour de consultation à l'autre (Hasson *et al.*, 2000). Des études ayant demandé jusqu'à 25 tours de consultation ont été rapportées (Couper, 1984). Trop peu de tours peut produire des résultats invalides ou insignifiants, alors qu'un nombre élevé de tours entraînera en revanche des difficultés de participation (Campbell *et al.*, 1999). Il est recommandé de conduire deux ou trois tours de consultation, bien que les preuves soutenant cette recommandation soient très minces (Goodman, 1987 ; Keeney *et al.*, 2006 ; Linstone et Turoff, 1975 ; Powell, 2003 ; Steurer, 2011).

1.1.5. Évaluation finale et publication

Lorsqu'un consensus est atteint, il demeure important de transmettre un résumé des résultats aux participants et de leur offrir l'occasion de les commenter. Si les commentaires le requièrent, de nouveaux tours de consultation pourront alors être effectués (Cole *et al.*, 2013). Il faut noter, en définitive, que l'atteinte d'un consensus ne signifie pas que la question de recherche a été résolue une fois pour toutes. L'utilisation de la méthode Delphi permet toutefois de structurer la discussion et de définir les zones de débat restantes (Hasson *et al.*, 2000 ; Jones et Hunter, 1995).

1.2. Limites de la méthode Delphi

Les principales sources de difficultés (Cole *et al.*; Hasson *et al.*, 2000; Linstone et Turoff, 1975; Powell, 2003) relatives à la méthode Delphi sont: 1) le risque d'imposer la perspective des chercheurs aux participants; 2) le risque d'obtenir un consensus sur «le plus petit dénominateur commun»; 3) la fausse présomption que la méthode peut remplacer toutes les formes de communication; 4) le fait de sous-estimer les exigences de la méthode pour les participants, dont le temps investi doit être compensé; 5) les difficultés de communication liées aux différences culturelles et langagières; 6) la sursimplification du problème à l'étude; 7) l'utilisation d'une mauvaise approche pour synthétiser, interpréter et présenter les réponses des participants; 8) l'ignorance des désaccords, de telle sorte que les participants en désaccord quittent le processus et que le consensus obtenu est artificiel.

La méthode Delphi originale a été l'objet de nombreuses modifications, notamment: 1) en utilisant des questions fermées plutôt qu'ouvertes dès le premier tour; 2) en augmentant le nombre d'experts; 3) en recrutant non pas des experts au sens habituel du terme, mais plutôt des personnes possédant de l'information pertinente sur la question de recherche (p. ex., des patients plutôt que des professionnels de la santé); 4) en modifiant les méthodes de rappels (téléphoniques, postaux, etc.); 5) en réunissant les participants, tout en maintenant la confidentialité des réponses par l'utilisation de télévoteurs (dans la version originale de la méthode Delphi, les experts ne se rencontrent jamais [Couper, 1984]); 6) en organisant une rencontre des participants en cours de route ou à la fin, pour apporter certaines informations ou clarifier certains éléments – dans ce cas, l'animation et l'environnement doivent être soigneusement préparés (Reagan-Cirincione et Rohrbaugh, 1992); et 7) en limitant le retour d'information aux participants (Couper, 1984). Ainsi, la plupart des chercheurs utilisent aujourd'hui une méthode Delphi dite «modifiée» d'une façon ou d'une autre (Hasson *et al.*, 2000; Keeney *et al.*, 2006; Linstone et Turoff, 1975; Mullen, 2003) et l'affublent souvent de différents qualificatifs (p. ex., «Real-time *Delphi*», «*Decision Delphi*», «Fuzzy Delphi»). Ces qualificatifs, nombreux, concernent parfois le type d'application, la définition du consensus ou encore d'autres éléments de la méthode, mais ne reflètent pas une organisation systématique (Mullen, 2003).

1.3. Logiciels et autre matériel nécessaires pour utiliser la méthode Delphi

La méthode Delphi est somme toute relativement simple et, comme il a été dit plus haut, était conduite à l'origine par questionnaires postaux. De nos jours, l'utilisation de questionnaires Web requiert la maîtrise de logiciels

de conception de questionnaires en ligne, tels que SurveyMonkey, LimeSurvey, Google Docs, JotForm, FluidSurveys, FreeOnlineSurveys, Qualtrics et Microsoft Word (pour construire des questionnaires en documents joints à un message courriel) (Cole *et al.*, 2013; Jamal *et al.*, 2011). Des outils spécifiques à la méthode Delphi existent également, par exemple un logiciel développé par Deshpande et ses collaborateurs à l'Université Yale, qui calcule automatiquement l'accord selon la RAND/UCLA Appropriateness Method (Deshpande et Shiffman, 2003; Deshpande, Shiffman et Nadkarni, 2005). Si ce genre d'outils est retenu, il sera important de s'assurer que la fidélité et la validité des questionnaires sont comparables à celles de questionnaires postaux, que la sécurité des données est préservée, que la présentation visuelle est simple et de bonne qualité, que l'interface est facile à utiliser avec des moyens techniques courants, et que les temps de sauvegarde et de téléchargement sont réduits au minimum afin que les participants ne se découragent pas (Cole *et al.*, 2013; Holloway, 2012; Jamal *et al.*, 2011). Si les questionnaires Web sont utilisés, la vérification préalable des compétences informatiques des experts sera très importante et fera peut-être même partie des critères de sélection (Gill *et al.*, 2013). L'analyse des contenus qualitatifs générés principalement lors du premier tour pourra nécessiter l'utilisation d'un logiciel spécialisé comme NVivo (Bergin, 2011) ou Ethnograph (Pateman, 1998).

La gestion des données quantitatives est plutôt simple et pourra être faite avec le logiciel Microsoft Excel ou un autre tabulateur. Il s'agira alors de compiler des statistiques (mesures de tendance centrale: moyenne, médiane ou mode – mesures de dispersion: étendue, écart-type, écart intercentile ou intervalle interquartile – mesures d'accord: pourcentage d'accord, coefficient Kappa, coefficient Kappa pondéré) sur les niveaux d'accord des participants (Hasson *et al.*, 2000).

La méthode Delphi a été utilisée à de nombreuses reprises dans le domaine de la santé pour atteindre des consensus, sur des sujets aussi diversifiés que les besoins des personnes âgées (Clark et Cochran, 1972), le choix d'indicateurs de performance des systèmes de santé (Boulkedid *et al.*, 2011; Burnam, 2005), la pathologie médicale (Hill et Goodale, 1981), la qualité des soins palliatifs (Morita *et al.*, 2005) et les soins psychiatriques (Kumaran, Hansen et Rowe, 1976), pour n'en nommer que quelques-uns. Dans la section qui suit, nous présenterons un exemple d'application de la méthode Delphi en santé mentale au travail.

2. EXEMPLE D'APPLICATION DE LA MÉTHODE DELPHI EN SANTÉ MENTALE AU TRAVAIL

Cette partie sera consacrée à l'illustration pratique de l'utilisation de la méthode Delphi modifiée (Tremblay-Boudreault, 2012) pour concevoir un questionnaire sur la charge de travail mentale (CTM), destiné aux personnes qui sont en processus de retour au travail à la suite d'un trouble mental transitoire (TMT). Jusque-là, aucun outil n'était disponible à cette fin (voir le chapitre 6 de cet ouvrage pour plus de détails sur l'outil). La méthode Delphi a permis d'évaluer la validité apparente (pertinence), la clarté et la formulation des items d'un questionnaire en phase de conception, et ce, en recourant à des experts en réadaptation au travail qui œuvrent auprès de cette clientèle.

2.1. Justification de la méthode Delphi pour cette application

L'une des utilités de la méthode Delphi est l'accompagnement à la conception d'outils cliniques (Gaskin, O'Brien et Hardy, 2003). Dans l'exemple qui sera illustré ici, les chercheurs ont utilisé une adaptation de la méthode publiée par Minkman *et al.* (2009). Comme l'adaptation effectuée par ces auteurs favorise le courriel comme mode d'échanges, elle a été retenue, car elle permet de joindre rapidement et à moindre coût les experts québécois en réadaptation au travail en santé mentale. En effet, les experts reconnus dans ce domaine sont encore peu nombreux et ils sont répartis dans différentes régions québécoises. De plus, dans la présente étude, étant donné qu'il n'y avait pas de consensus quant à la définition même de la CMT (Neill, 2011 ; Wickens, 2008), la méthode Delphi représentait un choix approprié en complément à l'analyse de concepts, puisque cette méthode permet d'approfondir et d'évaluer la validité apparente des items conçus à partir des résultats de l'analyse de concepts (voir le chapitre 6 de cet ouvrage). L'utilisation des questionnaires en format Microsoft Word a permis d'améliorer les définitions des éléments identifiés lors de l'analyse de concepts portant sur la CTM, de compléter ces dernières et de réduire le nombre d'éléments. Dans les prochains paragraphes, nous suivrons la logique des phases préconisée dans la première partie de ce chapitre pour rendre compte de cette application de la méthode Delphi.

2.2. Procédure de collecte, d'analyse et d'interprétation des résultats

2.2.1. Mise sur pied d'un panel d'individus informés (« experts »)

Des experts en réadaptation au travail impliqués en recherche ou en clinique auprès de travailleurs en processus de retour au travail à la suite d'un TMT ont été recrutés par la méthode « boule de neige » afin de commenter différentes versions du questionnaire de la CMT en cours de conception. Pour chacune des catégories d'experts ciblées (chercheurs, ergonomes, ergothérapeutes, neuropsychologues, psychiatres, psychologues), une personne était d'abord identifiée par l'équipe de recherche comme étant un expert dans le domaine, en fonction des critères d'inclusion suivants : 1) posséder une expérience en réadaptation au travail en santé mentale d'au moins deux ans ; 2) avoir une expérience clinique minimale de trois ans ; 3) lire et écrire le français ; et 4) être disponible pour la période visée par l'étude. Par la suite, chaque expert contacté était invité à donner la référence d'un pair qu'il considérait comme un expert répondant aux mêmes critères, et ce, jusqu'à l'obtention de deux à trois experts par catégorie. Ainsi, trois chercheurs et treize cliniciens (trois ergonomes, trois ergothérapeutes, trois neuropsychologues, deux psychiatres et deux psychologues) ont pris part à la consultation.

2.2.2. Exploration du sujet et retour des questionnaires

Étant donné que les thèmes et le premier jet des items du questionnaire ont découlé d'une analyse de concepts (voir le chapitre 6 de cet ouvrage), le premier tour de consultation du Delphi portait sur l'évaluation de la pertinence et de la formulation des items de la version 1 du questionnaire de la charge de travail mentale (QCTM-1). Les experts du panel ont reçu un courriel personnalisé auquel était joint un fichier d'évaluation du QCTM-1 (fichier Microsoft Word en mode formulaire), composé de questions fermées et de questions ouvertes. Les questions fermées évaluaient la validité apparente de chaque item à partir d'une échelle de pertinence à neuf niveaux (de 1 = pas du tout pertinent à 9 = tout à fait pertinent). Les questions ouvertes étaient plutôt destinées à recueillir des commentaires sur la formulation des items. Les figures 13.1 et 13.2 présentent respectivement des extraits des questions fermées et ouvertes soumises aux experts. Chaque expert disposait d'une période de dix jours pour réaliser son évaluation du QCTM-1 et la retourner par courriel. Au besoin, un courriel de rappel était envoyé sept jours après l'envoi initial, alors qu'un courriel de remerciement était envoyé après s'être assuré que toutes les sections avaient été remplies correctement par l'expert. En cas de données manquantes, l'expert était relancé par courriel pour obtenir les informations pertinentes.

Figure 13.1
Extrait d'une question fermée évaluant la validité apparente des items du QCTM-1

3.27. Vous avez eu des difficultés à prendre des décisions.

Presque jamais	Parfois	Souvent	Presque toujours
☐	☐	☐	☐

3.28. Vous avez manqué d'énergie.

Presque jamais	Parfois	Souvent	Presque toujours
☐	☐	☐	☐

3.29. Vous vous êtes senti(e) négatif(ve) envers les autres.

Presque jamais	Parfois	Souvent	Presque toujours
☐	☐	☐	☐

Item 3.27 :

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☐	☐	☐	☐	☐

Item 3.28 :

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☐	☐	☐	☐	☐

Item 3.29 :

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☐	☐	☐	☐	☐

LEGENDE POUR LA COTATION :
1 = pas du tout pertinent

9 = tout à fait pertinent

Figure 13.2.
Extrait des questions ouvertes évaluant la forme des items du QCTM-1

COMMENTAIRES GÉNÉRAUX POUR L'ÉVALUATION DE LA SECTION 3 :

Veuillez taper vos commentaires ici :

Vous pouvez vous référer aux critères suivants :

- La syntaxe et le vocabulaire sont-ils faciles à comprendre?
- Le texte est-il clair (absence d'ambiguïté) à la lecture des instructions?
- Des erreurs de grammaire sont-elles présentes?
- La longueur des consignes et des énoncés est-elle adéquate?
- Certains mots ou expressions pourraient-ils causer un inconfort chez la personne?
- Les items couvrent-ils bien l'aspect de la charge de travail mentale traité dans la section?
- Certains énoncés comprennent-ils trop d'éléments dans leur formulation et pour lesquels il serait judicieux de les séparer?
- Certains énoncés orientent-ils la direction de la réponse par leur formulation?
- Certains énoncés suscitent-ils la désirabilité sociale?
- L'échelle de réponse est-elle facile d'utilisation?
- D'autres commentaires vous viennent-ils à l'esprit?

2.2.3. Synthèse des positions individuelles et définition des zones d'accord et de désaccord

Analyse des données quantitatives

La définition de consensus de la RAND/UCLA Appropriateness Method (RAM) (Fitch *et al.*, 2001) a été utilisée pour analyser les données quantitatives sur la pertinence des items. Pour un item donné, la valeur de la médiane représentait la tendance centrale de l'opinion des experts quant à la pertinence de l'item. Trois options étaient donc possibles pour la pertinence de chaque item : non pertinent (médiane entre 1 et 3), pertinence incertaine (médiane entre 4 et 6) et pertinent (médiane entre 7 et 9). Quant au consensus des experts, il était déterminé à partir de la comparaison de deux mesures de dispersion (l'écart intercentile et l'écart intercentile ajusté pour la symétrie). Un consensus (aussi appelé accord des experts dans le tableau 13.1) était considéré comme atteint pour un item si l'écart intercentile pour cet item était inférieur à l'écart intercentile ajusté pour la symétrie pour ce même item, alors que l'inverse indiquait plutôt une absence de consensus (aussi appelée désaccord des experts dans le tableau 13.1). Ainsi, la pertinence, conjuguée au consensus, déterminait ce qu'il adviendrait des items. Le tableau 13.1 présente l'intégration de l'interprétation de la méthode RAM appliquée à l'évaluation de la pertinence pour les items du QCTM-1 selon le niveau de consensus atteint.

Tableau 13.1.
Interprétation des analyses de la cotation des experts du Delphi

Niveau de pertinence (médiane du groupe d'experts)	Accord des experts (EI < EIAS)	Désaccord des experts (EI > EIAS)
7-9 = pertinent	Accord de pertinence[a]	Désaccord de pertinence[b]
4-6 = pertinence incertaine	Accord de pertinence incertaine[b]	Désaccord de pertinence incertaine[b]
1-3 = non pertinent	Accord de non-pertinence[c]	Désaccord de non-pertinence[b]

Légende : EI = écart intercentile ; EIAS = écart intercentile ajusté pour la symétrie.
a : item conservé d'emblée b : item soumis au second tour de consultation c : item rejeté d'emblée.

Comme présenté, les combinaisons de pertinence et de consensus indiquent les possibilités suivantes : *a*) un accord de pertinence menant à la conservation d'emblée d'un item, soit lorsque les experts sont d'accord qu'un item est pertinent (médiane entre 7 et 9) ; *b*) un accord de pertinence incertaine ou toute situation de désaccord conduisant à la soumission de l'item au second tour de consultation, soit lorsque les experts sont

d'accord qu'un item a une pertinence incertaine (médiane entre 4 et 6), soit lorsque les experts sont en désaccord peu importe la cote médiane de pertinence obtenue ; et *c*) un accord de non-pertinence impliquant le rejet d'emblée d'un item, soit lorsque les experts sont d'accord qu'un item n'est pas pertinent (médiane entre 1 et 3).

Pour permettre de traiter efficacement l'ensemble des données quantitatives recueillies par les experts, une banque de données a été conçue dans un fichier Microsoft Excel, dont un extrait est présenté à la figure 13.3. Ainsi, la première colonne correspondait au numéro d'item du QCTM-1 évalué par les experts, alors qu'une colonne était destinée à la cotation donnée par chaque expert. En saisissant les cotations de chaque expert, les formules automatiques insérées dans les colonnes suivantes permettaient de calculer les statistiques descriptives (médiane, 30e centile, 70e centile, écart intercentile, indice [ou index] d'asymétrie et écart intercentile ajusté pour la symétrie) requises pour l'interprétation des résultats (colonne « accord ou désaccord » et colonne « opinion des experts »). Il est à noter que le fichier Microsoft Excel avait été programmé pour que les valeurs « accord, désaccord, non pertinent, pertinence incertaine ou pertinent » apparaissent automatiquement pour chaque item. Pour une personne possédant des connaissances de base du logiciel Microsoft Excel, il est facile d'insérer automatiquement les formules ou les opérations logiques requises en les tapant directement ou en utilisant les formules déjà programmées dans le fichier. Avec ce type de fichier, il devient facile et rapide de présenter la cotation de l'expert par rapport à celle du groupe (p. ex., médiane et étendue des données). Pour le lecteur intéressé, les formules pertinentes et un exemple de calcul complet sont présentés à l'annexe 13.1.

Au terme du premier tour de consultation, un consensus a été atteint pour la majorité des items. La plupart ont été jugés pertinents par les experts (p. ex., les items 3.1 et 3.28 à la figure 13.3), alors que deux items ont été jugés incertains (p. ex., les items 3.6 et 3.14 à la figure 13.3) et aucun n'a été jugé non pertinent. Un désaccord a été obtenu pour un seul item (l'item 3.14).

Analyse des données qualitatives et leur intégration aux données quantitatives

Les données qualitatives recueillies à partir des questions ouvertes à la fin de chaque section du QCTM-1 ont fait l'objet d'une synthèse des commentaires inspirée d'une analyse de contenu. Les données qualitatives ont été intégrées aux données quantitatives afin de prendre des décisions sur la conservation, le rejet et l'ajout d'items, ainsi que sur la formulation des items et des consignes.

Figure 13.3.
Extrait de la banque de données quantitatives construite dans le logiciel Microsoft Excel (premier tour de consultation)

Numéro item	Expert 1	Expert 2	...	Médiane	Centile 30	Centile 70	ÉI	Index	ÉIAS	ACCORD	OPINION DES EXPERTS
Section 3											
Item 3.1	9	9	...	8,50	8,00	9,00	1,00	-3,50	7,60	Accord	de pertinence
Item 3.6	9	9	...	6,00	5,50	8,50	3,00	-2,00	5,35	Accord	de pertinence incertaine
Item 3.14	9	9	...	6,00	3,00	8,00	5,00	-0,50	3,10	Désaccord	de pertinence incertaine
Item 3.28	9	9		9,00	8,00	9,00	1,00	-3,50	7,60	Accord	de pertinence

Notes : ÉI = écart intercentile ; Index = index d'asymétrie ; ÉIAS = écart intercentile ajusté pour la symétrie.

Un nouvel onglet du fichier Microsoft Excel a été conçu pour permettre de faire la synthèse des données qualitatives recueillies sur la qualité des items du QCTM-1, et ce, basé sur les thématiques des questions ouvertes. Un extrait est présenté à la figure 13.4. Ainsi, les deux premières colonnes étaient attribuées à la sous-section du questionnaire et à la thématique du commentaire. L'extrait textuel du commentaire correspondait au contenu prévu dans la troisième colonne, alors que les colonnes suivantes servaient pour identifier l'expert ayant mentionné le commentaire.

Figure 13.4.
Extrait de la banque de données qualitatives dans Microsoft Excel (second tour de consultation)

Section 3	Thématique du commentaire	Extrait du commentaire	Expert 1	Expert 2	...	Expert 16
	Arrimage entre "en raison", la consigne et les énoncés.	Section 3: le terme "en raison de" ne s'harmonise souvent pas avec le reste de l'énoncé dans les sous quesitons. Il serait préférable de l'ajuster selon les divers items. Ou encore clarifier les explicaitons et le sens de cette question dans le carré gris au début de cette section.		1		
	Clarification d'un item	3.14, formulation un peu complexe,				1
	Lien entre effort et un item	3.12 le lien entre le terme effort et confiance en soi est difficile à suivre: comment avoir plus ou moins confiance en moi exige plus ou moins d'effort? Pas instinctif comme lien!		1		

Dans le cadre de cette étude, les commentaires recueillis à partir des questions ouvertes (données qualitatives) ont été intégrés aux cotations quantitatives afin d'approfondir les données quantitatives (p. ex., « Que signifie exactement la faible cotation de pertinence pour cet item ? Comment pouvons-nous l'améliorer ? ») et de produire la seconde version du QCTM (QCTM-2). Les données qualitatives ont contribué à l'amélioration de la clarté des items et des consignes des différentes sections du questionnaire.

En effet, à la suggestion des experts, les échelles de cotation de la section 2 du QCTM-1 ont été modifiées afin de rendre leur utilisation plus intuitive. Un autre exemple d'intégration des données qualitatives concerne le retrait d'items. Quoique la plupart des items d'une sous-section du QCTM aient été jugés pertinents par les experts, des commentaires indiquant que certains items semblaient redondants ont également été considérés. À cet effet, les chercheurs ont décidé d'éliminer d'emblée deux items d'une même sous-catégorie qui avaient obtenu les cotations de pertinence les plus basses (pertinence entre 4 et 6). Par conséquent, il n'a pas été nécessaire de soumettre à nouveau ces items au second tour de consultation, comme il aurait été prévu si seuls les résultats des données quantitatives avaient été pris en compte. Ceci n'est qu'un survol des possibilités d'intégration des données qualitatives aux données quantitatives dans le contexte de la conception d'un questionnaire mesurant la CTM. Les données qualitatives obtenues dans le cadre d'une étude Delphi apportent sans contredit une profondeur et une richesse non négligeables aux données quantitatives.

2.2.4. Exploration des désaccords et retour aux participants

Comme un consensus avait été atteint pour l'ensemble des items retenus au terme du premier tour de consultation et que les suggestions d'amélioration ciblaient généralement les mêmes thématiques (p. ex., l'échelle de cotation de la section 2), un second tour a été réalisé principalement afin d'évaluer plus en profondeur le QCTM-2. Plus précisément, le second tour a porté sur la clarté de l'ensemble des items, ainsi que sur la pertinence des nouveaux items proposés à la suite des commentaires recueillis avec les questions ouvertes au tour précédent. La clarté a été quantifiée à partir d'une échelle de 9 points (de 1 = pas du tout clair à 9 = tout à fait clair). Une médiane de 7 obtenue par consensus était considérée comme l'indicateur d'une clarté satisfaisante. Toute autre combinaison entraînait une reformulation des items visés. La procédure employée a été la même pour les deux tours de consultation.

2.2.5. Évaluation finale et publication

Au terme des deux tours de consultation, une stabilité des commentaires a été atteinte. Le consensus a été atteint dès le second tour pour les nouveaux items proposés dans le QCTM-2 et la clarté a été jugée satisfaisante par les experts (médiane supérieure ou égale à 7) pour la majorité des items. Les zones de débats restantes concernaient la difficulté d'utilisation de l'échelle de cotation pour une section précise du QCTM-2, ainsi que le manque de considération accordée aux items positifs de la charge de travail mentale

pour le travailleur. Les experts soulignaient les problématiques persistantes du questionnaire sans toutefois être en mesure de proposer une solution pour rendre l'échelle de cotation plus claire ou pour intégrer la dimension positive de la CTM. Une version du QCTM-3 a été conçue à partir des commentaires recueillis au second tour de consultation. Toutefois, il était peu probable qu'un troisième tour de consultation permette d'avancer considérablement sur ces aspects puisqu'une plus grande profondeur de réflexion était requise. Il a été décidé d'aborder les zones de débat restantes lors de la prochaine étape du processus de conception avec une autre méthode que la méthode Delphi, soit des entrevues individuelles semi-structurées réalisées auprès de cinq travailleurs et de cinq cliniciens ayant été impliqués dans un processus de retour au travail à la suite d'un trouble mental transitoire. La méthode Delphi a donc contribué à préciser les items qui avaient été conçus à partir de l'analyse de concepts, ainsi qu'à cibler des zones de débats qui nécessitaient d'être approfondies par des entrevues individuelles dans une étape ultérieure du processus de conception du questionnaire. Le lecteur intéressé à en apprendre davantage sur le processus de conception du QCTM peut se référer aux travaux de Tremblay-Boudreault (2012).

CONCLUSION

Ce chapitre avait pour but d'exposer la méthode Delphi et d'en illustrer l'application dans un contexte de réadaptation en santé mentale. Nous avons tenté de démontrer que la méthode Delphi est une approche simple et efficace pour obtenir une réponse par consensus d'experts sur un problème qui ne peut être résolu par les méthodes de recherche qualitatives et quantitatives habituelles. Que ce soit pour accompagner le processus de conception d'un outil clinique, tel qu'illustré dans la seconde partie du chapitre, ou encore pour connaître l'opinion de professionnels de la santé sur les meilleures pratiques, la méthode Delphi permet l'intégration de données qualitatives aux données quantitatives dans un large éventail d'applications. Il s'agit d'une approche éprouvée et facile d'utilisation, qui constitue un outil de choix pour favoriser l'avancement des connaissances et répondre aux multiples enjeux rencontrés par les chercheurs et les cliniciens qui œuvrent auprès d'une clientèle présentant des troubles mentaux.

Annexe 13.1.
Illustration complète d'un exemple de calcul

Étape 1. Déterminer la cote médiane obtenue pour cet item	Considérons la distribution obtenue pour l'item 3.28 au premier tour de consultation (*n* = 16) : 2, 3, 5, 8, 8, 8, 8, 9, 9, 9, 9, 9, 9, 9, 9, 9. La première option pour déterminer la médiane est de repérer visuellement la valeur « milieu », après avoir ordonné les valeurs de cette distribution. Comme il s'agit d'un nombre pair de données, la médiane correspond à la valeur située entre la 8e et la 9e donnée, soit la valeur « 9 ». Il est à noter que si la 8e donnée et la 9e donnée avaient été respectivement de 6 et 7, la médiane aurait alors été à mi-chemin entre les deux, soit 6,5. La seconde option pour déterminer la médiane, plus efficace pour un grand nombre de données, consiste à utiliser une formule mathématique telle qu'illustrée dans les encadrés suivants.
Étape 2. Déterminer le 30e et le 70e centile de la distribution (respectivement nommé C_{30} et C_{70})	Pour déterminer ces valeurs, les formules mathématiques suivantes ont été employées. Il est à noter que comme la médiane correspond au 50e centile, alors les formules du 30e et du 70e centile sont une adaptation de la formule donnée pour la médiane. **Formules mathématiques:** **C_{70} = valeur du [X + (0,7*Y)]^{e} rang et C_{30} = valeur du [X + (0,3*Y)]^{e} rang** **Application des formules à l'item 3.28:** **C_{70} = valeur du [1 + (0,7*15)]^{e} rang et** **C_{30} = valeur du [1 + (0,3*15)]^{e} rang** où X et Y correspondent aux mêmes valeurs que celles présentées dans l'encadré sur l'application de la formule de la médiane. Donc: **C_{70}** = 1 + (0,7*15) = 11,5^{e} valeur = 9 **$C_{30 =}$** = 1 + (0,3*15) = 5,5^{e} valeur = 8 Donc: **C_{70}** = 1 + (0,7*15) = 11,5^{e} valeur = 9 **$C_{30 =}$** = 1 + (0,3*15) = 5,5^{e} valeur = 8
Étape 3. Calculer l'écart intercentile (entre le 70e et le 30e centile)	Pour déterminer cette valeur, la formule suivante a été employée. **Formules mathématiques:** **Écart intercentile = ÉI = $C_{70} - C_{30}$** **Application de la formules à l'item 3.28:** **Écart intercentile = ÉI = $C_{70} - C_{30}$** où **C_{70}** = 9 et **C_{30}** = 8 Donc: **ÉI** = 9 – 8 = 1

Étape 4. Calculer l'indice d'asymétrie	Pour déterminer la valeur de cet indice, la formule suivante a été employée. Cette valeur sera utile pour déterminer par la suite l'écart intercentile ajusté pour la symétrie. **Formule mathématique:** **Indice d'asymétrie = IA = | (Z – [(C_{30} + C_{70})/2]) |** **Formule mathématique appliquée à l'item 3.28:** **Indice d'asymétrie = IA** = | (5 – [(C_{30} + C_{70})/2]) | où **C_{30}** et **C_{70}** correspondent aux valeurs déterminées précédemment. **Z** = 5 (correspond à la valeur médiane obtenue sur l'échelle de cotation de 9 niveaux employée dans cette étude). Rappelons que les «|» au début et à la fin de la formule renvoient à la valeur absolue, c'est-à-dire que le signe n'importe pas. Donc: **IA** = | (5 – [(8 + 9)/2]) | = |- 3,5 | = 3,5
Étape 5. Calculer l'écart intercentile ajusté pour la symétrie	**Pour déterminer cette valeur, la formule suivante a été employée.** **Formule mathématique:** **Écart intercentile ajusté pour la symétrie = ÉIAS = 2,35 + (1,5*IA)** **Application de la formule à l'item 3.28:** **Écart intercentile ajusté pour la symétrie** = ÉIAS **= 2,35 + (1,5*IA)** où **IA** = 3,5 (voir l'encadré précédent). Donc: **ÉIAS** = 2,35 + (1,5*3,5) = 7,6
	Ainsi, pour l'item 3.28, comme l'écart intercentile ajusté pour la symétrie (EIAS = 7,6) était supérieur à l'écart intercentile (EI = 1), le consensus des experts a été atteint au premier tour de consultation. L'item a été conservé puisque la médiane obtenue pour la pertinence était de 9.

RÉFÉRENCES

Akins, R.B., H. Tolson et B.R. Cole (2005). « Stability of response characteristics of a Delphi panel : Application of bootstrap data expansion. [Validation Studies] », *BMC Med Res Methodol*, vol. 5, nº 37, <doi : 10.1186/1471-2288-5-37>.

Bantel, K. (1999). « Comprehensiveness of strategic planning : The importance of heterogeneity of a top team », *Psychol Rep*, vol. 73, p. 35-49.

Beech, B. (1999). « Go the extra mile-use the Delphi Technique », *J Nurs Manag*, vol. 7, nº 5, p. 281-288.

Bergin, M. (2011). « NVivo 8 and consistency in data analysis : Reflecting on the use of a qualitative data analysis program », *Nurse Res*, vol. 18, nº 3, p. 6-12.

Boulkedid, R., H. Abdoul, M. Loustau, O. Sibony et C. Alberti (2011). « Using and reporting the Delphi method for selecting healthcare quality indicators : A systematic review », *PLoS One*, vol. 6, nº 6, p. e20476.

Broad, W.J. (2006). *The Oracle : Ancient Delphi and the Science Behind Its Lost Secrets*, New York, Penguin Books.

Burnam, A. (2005). « Commentary : Selecting performance measures by consensus : An appropriate extension of the Delphi method ? », *Psychiatr Serv*, vol. 56, nº 12, p. 1583.

Campbell, S.M., M. Hann, M.O. Roland, J.A. Quayle et P.G. Shekelle (1999). « The effect of panel membership and feedback on ratings in a two-round Delphi survey : Results of a randomized controlled trial », *Med Care*, vol. 37, nº 9, p. 964-968.

Clark, L.H. et S.W. Cochran (1972). « Needs of older Americans assessed by Delphi procedures », *J Gerontol*, vol. 27, nº 2, p. 275-278.

Clibbens, N., S. Walters et W. Baird (2012). « Delphi research : Issues raised by a pilot study », *Nurse Res*, vol. 19, nº 2, p. 37-44.

Cole, Z.D., H.M. Donohoe et M.L. Stellefson (2013). « Internet-based Delphi research : Case based discussion », *Environ Manage*, vol. 51, nº 3, p. 511-523.

Couper, M.R. (1984). « The Delphi technique : Characteristics and sequence model », *Adv Nurs Sci*, vol. 7, nº 1, p. 72-77.

Dalkey, N et O. Helmer (1963). « An experimental application of the Delphi method to the use of experts », *Management Science*, vol. 9, nº 3, p. 458-467.

Deshpande, A.M. et R.N. Shiffman (2003). « Delphi rating on the Internet », *AMIA Annu Symp Proc*, p.828

Deshpande, A.M., R.N. Shiffman et P.M. Nadkarni (2005). « Metadata-driven Delphi rating on the Internet », *Comput Methods Programs Biomed*, vol. 77, nº 1, p. 49-56.

Duffield, C. (1988). « The Delphi technique », *Aust J Adv Nurs*, vol. 6, nº 2, p. 41-45.

Duffield, C. (1993). « The Delphi technique : A comparison of results obtained using two expert panels », *Int J Nurs Stud*, vol. 30, nº 3, p. 227-237.

Fitch, K., J.B. Steven, M.S. Aguilar, B. Burnand, J.R. LaCalle, P. Lazaro et J.P. Kahan (2001). *The RAND/UCLA Appropriateness Method User's Manual*, Santa Monica, RAND Corporation

GASKIN, C.J., A.P. O'BRIEN et D.J. HARDY (2003). « The development of a professional practice audit questionnaire for mental health nursing in Aotearoa/New Zealand », *International Journal of Mental Health Nursing*, vol. 12, nº 4, p. 259-270, <doi : 298 [pii]>.

GILL, F.J., G.D. LESLIE, C. GRECH et J.M. LATOUR (2013). « Using a web-based survey tool to undertake a Delphi study : Application for nurse education research », *Nurse Educ Today*, vol. 33, p. 1322-1328.

GOODMAN, C.M. (1987). « The Delphi technique : A critique », *J Adv Nurs*, vol. 12, nº 6, p. 729-734.

HASSON, F., S. KEENEY et H. MCKENNA (2000). « Research guidelines for the Delphi survey technique », *J Adv Nurs*, vol. 32, nº 4, p. 1008-1015.

HILL, R.B. et F. GOODALE (1981). « The Delphi predictions of pathology chairmen : A six-year retrospective view », *J Med Educ*, vol. 56, nº 7, p. 537-546.

HOLEY, E.A., J.L. FEELEY, J. DIXON et V.J. WHITTAKER (2007). « An exploration of the use of simple statistics to measure consensus and stability in Delphi studies », *BMC Med Res Methodol*, vol. 7, p. 52.

HOLLOWAY, K. (2012). « Doing the E-Delphi : Using online survey tools », *Comput Inform Nurs*, vol. 30, nº 7, p. 347-350.

JAMAL, F., C. THOER, E. DUCHARME et F. BESNER (2011). « Les méthodes de recherche en ligne en santé des populations : une adaptation des méthodes de recherche classique ? », *Carnet synthèse*, nº 10, p. 12.

JONES, J. et D. HUNTER (1995). « Consensus methods for medical and health services research », *BMJ*, vol. 311, nº 7001, p. 376-380.

KEENEY, S., F. HASSON et H. MCKENNA (2006). « Consulting the oracle : Ten lessons from using the Delphi technique in nursing research. [Review] », *Journal of Advanced Nursing*, vol. 53, nº 2, p. 205-212, <doi : 10.1111/j.1365-2648.2006.03716.x>.

KUMARAN, K., R.C. HANSEN et M. ROWE (1976). « The Delphi technique in a psychiatric hospital », *Dimens Health Serv*, vol. 53, nº 8, p. 32-34.

LINSTONE, H.A. et M. TUROFF (1975). *The Delphi Method : Technique and Applications*, Londres, Addison-Wesley.

MCMURRAY, A.R. (1994). « Three decision-making aids : Brainstorming, nominal group, and Delphi technique », *Journal Nursing Staff Dev*, vol. 10, nº 2, p. 62-65.

MINKMAN, M., K. AHAUS, I. FABBRICOTTI, U. NABITZ et R. HUIJSMAN (2009). « A quality management model for integrated care : Results of a Delphi and Concept Mapping study », *International Journal of Quality in Health Care*, vol. 21, nº 1, p. 66-75, <doi : 10.1093/intqhc/mzn048 [pii]>.

MORITA, T., S. BITO, Y. KURIHARA et Y. UCHITOMI (2005). « Development of a clinical guideline for palliative sedation therapy using the Delphi method. [Research Support, Non-U.S. Gov't] », *J Palliat Med*, vol. 8, nº 4, p. 716-729, <doi : 10.1089/jpm.2005.8.716>.

MULLEN, P.M. (2003). « Delphi : Myths and reality », *J Health Organ Manag*, vol. 17, nº 1, p. 37-52.

Neill, D. (2011). « Nursing workload and the changing health care environment : A review of the literature », *Admininistrative Issues Journal*, vol. 1, nº 2, p. 132-143.

Pateman, B. (1998). « Computer-aided qualitative data analysis : The value of NUD*IST and other programs », *Nurs Res*, vol. 5, nº 3, p. 77-89.

Powell, C. (2003). « The Delphi technique : Myths and realities », *Journal of Advanced Nursing*, vol. 41, nº 4, p. 376-382.

Reagan-Cirincione, P. et J. Rohrbaugh (1992). « Decision conferencing : A unique approach to the behavioral aggregation of expert judgment Part II », dans G. Wright et F. Bolger (dir.), *Expertise and Decision Support*, New York, Springer, p. 181-201.

Steurer, J. (2011). « The Delphi method : An efficient procedure to generate knowledge », *Skeletal Radiol*, vol. 40, nº 8, p. 959-961.

Tremblay-Boudreault, V. (2012). *Conception d'un outil de mesure de la charge de travail mentale dans le processus de retour au travail à la suite d'un trouble mental transitoire ayant engendré des incapacités prolongées*, Mémoire de maîtrise, Sherbrooke, Université de Sherbrooke.

Van de Ven, A. et A. Delbecq (1972). « The nominal group as a research instrument for exploratory health studies », *Am J Public Health*, vol. 62, p. 337-342.

Wickens, C.D. (2008). « Multiple resources and mental workload », *Human Factors*, vol. 50, nº 3, p. 449-455.

CHAPITRE 14

TRIAGE – UNE TECHNIQUE STRUCTURÉE SOLLICITANT L'OPINION D'EXPERTS EN VUE D'ATTEINDRE UN CONSENSUS

Un exemple d'utilisation dans une recherche visant l'adaptation d'un questionnaire autoadministré pour une clientèle en incapacité prolongée au travail

Valérie Albert
Marie-José Durand
Geneviève Pepin

FORCES

- La structure de TRIAGE permet d'obtenir un ou des consensus qui tiennent compte de l'ensemble des opinions des experts par rapport à une question particulière.
- Les discussions entre les experts permettent l'émergence de nouvelles données.
- Le support visuel favorise et maintient l'implication active des experts, tout en facilitant la gestion des discussions et l'obtention du ou des consensus.

LIMITES

- La qualité des résultats dépend des habiletés de l'animateur et du choix des experts.
- L'absence d'un ou de plusieurs experts lors de la production collective, liée soit à l'éloignement géographique ou encore à la disponibilité de ceux-ci, peut affecter significativement le processus de tri.
- Certains éléments qui créent des affrontements entre les experts sans faire consensus sont mis de côté afin de poursuivre les discussions, mais pourraient être importants pour le thème à l'étude.

Que ce soit pour la conception d'un outil de mesure, l'approfondissement d'un thème particulier ou le choix de priorités d'intervention, il est fréquent d'avoir recours à l'opinion d'un groupe d'experts en recherche, particulièrement lorsque peu de données sont disponibles dans les écrits sur un sujet particulier (Alain et Dessureault, 2009). Parmi les techniques de groupe axées sur l'atteinte d'un consensus ou la prise de décision, la technique du groupe nominal et la méthode Delphi sont probablement les plus répandues, alors que la technique TRIAGE est moins utilisée en recherche.

Brièvement, la technique du groupe nominal regroupe en une rencontre environ six à dix participants, qui doivent fournir un certain nombre de réponses à une question soumise à l'ensemble du groupe, d'abord par écrit, puis en proposant une réponse à la fois à tour de rôle (Delbecq, Van de Ven et Gustafson, 1975). Les interactions ne sont permises qu'avec l'animateur, de sorte que la compréhension de chaque réponse ou le point de vue de l'ensemble des participants demeure à un niveau superficiel. Toutes les réponses fournies sont inscrites sur un support visuel et doivent par la suite être mises en ordre d'importance par chacun des participants. Le résultat du vote correspond au consensus du groupe, ce qui rend les résultats rapidement disponibles. Cependant, le consensus basé sur une mise en rang peut entraîner le rejet de certains éléments d'intérêt qui sont moins populaires.

Pour sa part, la méthode Delphi (voir le chapitre 13 de cet ouvrage) s'effectue entièrement à distance auprès de cinq à quinze experts (Alain et Dessureault, 2009 ; Delbecq *et al.*, 1975). Des rondes de consultation sur un sujet précis sont effectuées par courriel ou envoi postal. L'analyse des réponses des experts suivant la première ronde de consultation permet aux chercheurs de faire ressortir les convergences et les divergences d'opinions entre les experts, qui leur sont retournées lors des rondes subséquentes jusqu'à l'obtention de consensus, ce qui peut s'étaler sur plusieurs mois. La longueur du processus de consultation et la possibilité de perdre des experts d'une ronde à l'autre sont d'ailleurs des limites de cette méthode.

Moins connue, la technique de recherche d'information par l'animation d'un groupe d'experts (TRIAGE) fait aussi partie des méthodes facilitant la prise de décision au sein d'un groupe. Elle a été élaborée de façon à pallier certaines limites des méthodes plus répandues telles que la technique du groupe nominal et la méthode Delphi (Gervais, Pepin et Carrière, 2000). TRIAGE a déjà été utilisée dans des études ayant pour objectifs de sélectionner des indicateurs d'efficacité dans le cadre d'une évaluation de programme (Gervais, 1996), d'élaborer un questionnaire évaluant la qualité des services de réadaptation chez une clientèle avec un traumatisme crânien (Swaine *et al.*, 2003), d'améliorer la compréhension de la contribution de gestionnaires pour favoriser la collaboration entre la famille et les

intervenants en réadaptation (Carrière *et al.*, 2004) ou encore d'adapter et d'implanter une intervention de prise de décision partagée aux réalités de la réadaptation au travail (Coutu *et al.*, 2011). Ce chapitre exposera d'abord les origines de TRIAGE et décrira chacune de ses étapes. Un exemple d'application de TRIAGE en recherche dans le domaine de la réadaptation sera aussi présenté afin d'illustrer sa mise en œuvre.

1. ORIGINE DE LA TECHNIQUE

TRIAGE s'inscrit dans une perspective constructiviste, qui admet la présence de plusieurs réalités dans une même situation qui sont construites en fonction des connaissances et des expériences des personnes (Gervais et Pepin, 2002). En adoptant cette perspective, les chercheurs renoncent ainsi à vouloir imposer leurs propres opinions ou convictions, pour laisser place à la délibération des experts fondée sur leur expérience (Leclerc, 1999). Il s'agit d'une technique inductive, où les interactions directes entre les participants permettent de stimuler leur réflexion, de confronter leurs points de vue et de faire émerger de nouvelles données. Issue du domaine de l'éducation, TRIAGE est une technique qui a été développée par Plante et Côté en 1993 dans le but de favoriser la prise de décision et l'atteinte de consensus à l'intérieur d'un groupe (Gervais et Pepin, 2002). Le consensus se situe sur un continuum entre la majorité et l'unanimité. Il est défini comme le processus coopératif selon lequel tous les membres d'un groupe arrivent à un accord pour appuyer une décision qui privilégie l'intérêt de l'ensemble, où la contribution de chaque membre est prise en considération et qu'un effort basé sur la bonne foi est effectué pour discuter de tous les points de vue légitimes (Dressler, 2006 ; St-Arnaud, 2008).

2. DESCRIPTION DE LA TECHNIQUE

Dans le cadre de TRIAGE, l'opinion des experts est sollicitée en deux temps. Une première réflexion individuelle à distance leur permet d'émettre des opinions sur une ou des questions à l'étude. Cette réflexion individuelle est suivie d'une rencontre réunissant tous les experts, où un tri de l'ensemble des opinions émises est progressivement effectué par consensus. Cependant, afin de bien comprendre la mise en œuvre de TRIAGE et les précautions que les chercheurs doivent prendre pour obtenir des résultats de qualité, cette technique sera présentée en quatre étapes : 1) préparation ; 2) production individuelle ; 3) compilation ; et 4) production collective.

2.1. Préparation

Pour les chercheurs, la préparation est la plus longue et la plus difficile des quatre étapes de TRIAGE. Il s'agit, d'une part, de préciser la ou les questions à l'étude et, d'autre part, de sélectionner les experts. La formulation des questions à l'étude est une étape cruciale pour permettre aux experts de fournir des réponses de qualité, qui seront sous forme de courts énoncés (indicateurs de performance, actions, items d'un questionnaire, etc.). Pour soutenir la réflexion des experts, de l'information standardisée leur est fournie dans un document préparatoire à lire avant de répondre aux questions. Le principal enjeu concernant le document préparatoire consiste à fournir suffisamment d'informations pour que l'expert soit mis en contexte et comprenne bien la technique, sans toutefois le submerger d'informations qui pourraient biaiser ses réponses ou émousser son intérêt en raison de la lourdeur de la tâche. Ce document comprend par exemple le but de l'étude, le modèle théorique ou le cadre conceptuel sur lequel reposent les questions à l'étude, la définition de concepts clés, une brève description de TRIAGE, des consignes et des exemples. Les exemples doivent par contre être utilisés avec une certaine prudence, pour ne pas éliminer d'emblée certaines réponses d'intérêt.

Par exemple, si le thème à l'étude est l'identification d'indicateurs de maintien au travail d'employés présentant un trouble mental, et que les exemples fournis sont le « niveau d'estime de soi en tant que travailleur » et la « qualité de la relation avec le superviseur », ces deux indicateurs ne seront vraisemblablement pas renommés par les experts et seraient potentiellement exclus des résultats de la recherche, et ce, malgré leur pertinence. Une stratégie alternative consiste à développer un formulaire de réponse clair et adapté aux besoins de l'étude, permettant d'orienter l'expert quant au format et au nombre de réponses à fournir, tout en réduisant le potentiel de données manquantes pour les chercheurs[1]. Certains ouvrages peuvent guider le développement du formulaire de réponse, tels que celui publié par Dillman, Smyth et Christian (2009). Dans le formulaire, il est particulièrement important de préciser le nombre maximal de réponses à fournir pour chaque question. En effet, comme il s'agit d'une technique axée sur l'obtention de consensus, chaque expert fournira d'emblée les réponses les plus importantes selon lui, permettant déjà un premier tri parmi l'ensemble des réponses possibles.

Avant de procéder à l'envoi des documents aux experts, un prétest qualitatif devrait être effectué. Il s'agit de soumettre la lettre de présentation, le document préparatoire et le formulaire de réponse à quelques

1. La figure 14.2 de ce chapitre présente un exemple de formulaire de réponse.

acteurs cibles répondant aux mêmes critères de sélection que les experts recrutés, afin d'estimer le temps requis pour terminer le travail, mais surtout pour s'assurer de la clarté du contenu, de l'adéquation du format et pour sonder le type de réponses obtenues (Gervais *et al.*, 2000). En effet, les documents peuvent apparaître clairs et précis pour les chercheurs qui sont familiers avec TRIAGE et les résultats attendus. Or l'expérience peut être très différente pour l'expert pour qui cette technique est inconnue, d'où l'importance d'être très clair dans les explications de TRIAGE, les instructions et les questions à l'étude. Les questionnements ou ambiguïtés relevés par les répondants du prétest devront évidemment être résolus avant l'envoi des documents aux experts.

Lors de la préparation, il est également profitable de réfléchir à l'avance à la durée estimée de la dernière étape de TRIAGE, soit la rencontre de groupe (production collective), puisque cette information permettra de prévoir la plage horaire, de réserver les locaux et d'en aviser les experts dès leur recrutement. L'estimation de la durée de la production collective doit tenir compte, entre autres, du nombre de questions à l'étude, du nombre maximal de réponses à fournir et du nombre d'experts dans le groupe. À titre indicatif, des rencontres de deux à trois heures permettent de traiter une quantité intéressante d'information, mais au-delà de cette période, le niveau de fatigue des experts et de l'animateur pourrait altérer les résultats produits.

La sélection des experts est tout aussi cruciale que la formulation des questions à l'étude. De façon générale, les experts recrutés pour TRIAGE, au nombre de six à douze, sont des informateurs clés, définis comme des personnes dont la crédibilité est reconnue par leurs pairs et capables de fournir une information valide sur les questions à l'étude (Gervais *et al.*, 2000). Un échantillonnage par choix raisonné s'avère la stratégie la plus appropriée pour assurer une grande diversité d'expériences et de connaissances, tout en formant un groupe relativement homogène (Geoffrion, 1997). En effet, l'homogénéité des experts est un élément clé pour favoriser la dynamique de groupe lors de la production collective. Par exemple, il n'est pas recommandé de réunir des utilisateurs de services et des cliniciens dans un même groupe, les uns pouvant être inhibés par la présence des autres (voir le chapitre 23 de cet ouvrage). Si l'expertise de plusieurs groupes d'acteurs est nécessaire, il est préférable d'effectuer deux ou plusieurs processus TRIAGE de façon parallèle. Une façon de combiner les résultats obtenus au terme de ces processus parallèles sera d'ailleurs abordée plus loin dans ce chapitre (section 2.4.4, « Stratégie pour fusionner les résultats issus de plusieurs groupes »).

Bien que présentées de façon séquentielle, l'élaboration des questions et la sélection des experts se réalisent de façon itérative lors de la phase de préparation. En effet, en circonscrivant les questions ou thèmes à l'étude, les chercheurs peuvent se positionner quant à quel type d'experts permettrait d'offrir les réponses les plus pertinentes, alors que le choix des experts peut à son tour influencer la formulation des questions et le type ou la quantité d'information à fournir. Lorsque les experts sont recrutés et que les documents préparatoires sont satisfaisants, la production individuelle peut être entreprise.

2.2. Production individuelle

La production individuelle consiste en une première réflexion à distance effectuée par chacun des experts sur les questions à l'étude (Gervais *et al.*, 2000). L'envoi à chaque expert de la lettre d'introduction, du document préparatoire et du formulaire de réponse est effectué de façon postale ou électronique. Un délai de deux à trois semaines est alloué aux experts pour remplir le formulaire de réponse et le retourner aux chercheurs. Ce délai doit être clairement indiqué dans la lettre d'introduction et précisé lors du recrutement. Un court rappel par courriel ou par la poste peut être envoyé quelques jours avant la fin de la période allouée, si certains experts n'ont pas retourné leur formulaire de réponse.

2.3. Compilation

La compilation consiste essentiellement en la transcription telle quelle des réponses figurant dans le formulaire rempli par chacun des experts sur des cartons individuels, dans l'objectif de traiter l'ensemble des réponses en groupe (Gervais *et al.*, 2000). La transcription telle quelle des réponses est absolument essentielle, puisque l'implication active des experts lors de la production collective dépend en grande partie du fait qu'ils voient leurs propres mots sur les cartons. De plus, toutes les réponses doivent être transcrites, même si les chercheurs jugent que certaines sont hors de propos ou que plusieurs experts ont fourni les mêmes réponses. En effet, le risque associé à toute forme de traitement de l'information par les chercheurs à l'étape de compilation est que l'expert ne retrouve plus sa contribution dans le processus de recherche, inhibant sa participation à la construction des consensus. Utiliser les propres mots de l'expert et l'ensemble de ses suggestions constitue ainsi une forme tangible de reconnaissance à l'égard de son expertise par rapport au thème à l'étude.

S'il y a plus d'une question à l'étude, une banque de cartons est constituée pour chaque question, puisque les questions seront traitées séparément au cours de la production collective. Lorsque toutes les réponses sont transcrites, les cartons de chaque banque sont classés en ordre alphabétique (premier mot de la réponse) et numérotés, afin de faciliter leur repérage par l'animateur et les experts lors de la production collective. Le classement des cartons de réponses par ordre alphabétique permet aussi de conserver la confidentialité de l'auteur, un élément important de la production collective qui sera abordé dans la prochaine section. L'expert reconnaîtra quand même ses réponses, puisque ce sont ses propres mots qui apparaîtront sur les cartons. Enfin, si plusieurs questions doivent être traitées lors d'une même rencontre, le temps à allouer pour chaque banque sera estimé en fonction du volume de cartons à traiter et de la diversité des réponses obtenues. En effet, une plus grande diversité dans les réponses pourrait laisser supposer une plus grande divergence d'opinions et l'obtention plus laborieuse de consensus. Il pourrait aussi s'agir, malheureusement, d'un indice que les questions à l'étude n'étaient pas suffisamment précises. Or il est difficile de faire marche arrière à l'étape de la compilation, d'où l'importance qu'il faut accorder à la définition des questions et au prétest lors de l'étape de préparation.

2.4. Production collective

La production collective réunit tous les experts en une même rencontre et vise à faire le tri des réponses obtenues lors de la production individuelle. L'objectif est de sélectionner par consensus les éléments les plus pertinents parmi tous ceux qui ont été émis. L'analyse des données est donc effectuée tout au long du tri, grâce au consensus des experts (Gervais et Pepin, 2002). À la fin de la production collective, les résultats sont immédiatement disponibles, ce qui en fait une technique économique en matière de temps et de ressources (Carrière *et al.*, 2004; Gervais *et al.*, 2000). Le tri s'appuie sur un support visuel qui consiste en un grand mur (ou tableau) séparé en sections (définies dans la prochaine section). Les réponses des experts inscrites sur des cartons sont déplacées progressivement par l'animateur, de section en section, en fonction de l'accord du groupe obtenu pour chaque déplacement de carton. Ces déplacements représentent d'ailleurs une preuve tangible de la progression du tri, ce qui est stimulant tant pour les experts que pour l'animateur (Gervais *et al.*, 2000).

2.4.1. Support visuel et procédure de tri

Afin d'avoir une vue d'ensemble sur le support visuel tout au long du tri, les participants doivent être placés en demi-lune autour d'une longue table faisant face à celui-ci. L'animateur doit quant à lui avoir suffisamment d'espace pour circuler aisément devant la table des experts, d'un côté à l'autre du support visuel, pour déplacer les cartons de section en section. La figure 14.1 illustre la disposition des grandes sections du support visuel de TRIAGE. Sur la figure, on retrouve la section MÉMOIRE COLLECTIVE, située à la gauche du support visuel, les sections REGROUPEMENT, POUBELLE et FRIGO au centre et les sections SÉLECTION et VETO à droite (termes définis plus bas). Sur le support visuel, il peut être très utile d'ajouter des images pour illustrer chacune des sections. En effet, les images facilitent l'assimilation des consignes de tri liées à chaque section et peuvent détendre l'atmosphère lorsqu'un consensus est difficile à atteindre (voir l'exemple du FRIGO ci-après). Les consignes reliées à chaque section à présenter aux participants avant d'amorcer le tri seront brièvement présentées.

Figure 14.1.
Support visuel utilisé lors de la production collective

La section MÉMOIRE COLLECTIVE contient initialement tous les cartons reliés à une question, collés à l'aide d'adhésif réutilisable pour être facilement déplaçables. Il s'agit de toutes les réponses émises par les experts lors de la production individuelle pour cette question (Gervais *et al.*, 2000). Dans un premier temps, l'animateur doit s'assurer que les experts comprennent chacun des énoncés inscrits sur les cartons de la MÉMOIRE COLLECTIVE. Au besoin, l'animateur raffinera le contenu s'il est en mesure de le faire ou l'auteur de l'énoncé ajoutera des précisions s'il le souhaite. Les cartons doivent ensuite être regroupés par les experts si certaines réponses présentent des similarités (section REGROUPEMENT) ou éliminés s'ils sont jugés non pertinents (section POUBELLE). À cet effet, les experts doivent être avisés que les suggestions qui atteignent la poubelle ne peuvent réintégrer les discussions : il s'agit de décisions définitives.

Lorsque des cartons sont regroupés, ils sont collés les uns par-dessus les autres dans la section REGROUPEMENT, en conservant sur le dessus le terme le plus juste selon les experts. De cette façon, si un questionnement par rapport à un carton émerge au fil des discussions ou si un regroupement est modifié, les cartons sont toujours disponibles. Un nouveau nom peut être donné à un regroupement si un thème plus général est formé (FUSIONNER). Une réponse contenant deux idées peut aussi être divisée en propositions distinctes (SCINDER) et de nouvelles propositions peuvent émerger des discussions entre les experts (AJOUTER). À cet effet, l'animateur doit prévoir des cartons vierges, de l'adhésif réutilisable supplémentaire et un crayon marqueur. En cas de désaccord entre les experts, l'animateur peut temporairement retirer un carton de la discussion en le plaçant dans la section FRIGO, pour le reprendre plus tard et ainsi éviter de prolonger des discussions momentanément infructueuses. Pour cette section, l'image du FRIGO est très éloquente : voir le « frigo » et mentionner que le carton sera « conservé au frais » pour y revenir plus tard permet de détendre l'atmosphère et de poursuivre le tri. Par contre, aucun carton ne peut demeurer dans cette section à la fin du tri. Si un désaccord insoluble persiste concernant un carton, il est placé dans la section VETO pour être soumis ultérieurement à d'autres experts, externes au groupe, afin de poursuivre le tri et de favoriser l'harmonie du groupe. Enfin, les cartons déplacés dans la section SÉLECTION constituent le résultat final, c'est-à-dire la ou les réponses à la question posée aux experts. Comme pour la section POUBELLE, les suggestions entrent dans la section SÉLECTION à la suite d'une décision définitive. Un nombre maximal de réponses admissibles dans cette section est déterminé à l'avance pour chaque question et exposé aux experts au début du tri.

2.4.2. Précautions à prendre lors de la production collective

Le bon déroulement de la production collective dépend directement de la qualité de l'animation. L'animateur devrait être compétent et expérimenté dans la gestion d'un groupe et maîtriser les techniques d'entrevue. Bien qu'il soit au second plan lors des discussions, l'animateur doit s'assurer que tous les experts participent, tout en gérant avec tact les participants dominants, d'où l'importance pour les chercheurs de désigner un animateur compétent. S'il s'agit d'une première expérience d'animation de TRIAGE, il est essentiel que l'animateur soit en mesure de bien expliquer le déroulement général de la production collective et les consignes reliées aux sections du support visuel. Faire une simulation de la procédure de tri avec un autre groupe ou être accompagné d'un membre observateur qui est familier avec TRIAGE représentent aussi des stratégies intéressantes pour s'assurer du bon déroulement du tri et de la qualité des résultats obtenus.

En entamant chaque banque de cartons ou question de l'étude, l'animateur sensibilise les participants au temps alloué pour la traiter et précise aussi que TRIAGE encourage une participation spontanée, sans tour de table en accordant la priorité à l'animateur en cas de confusion. Lorsqu'un accord semble émerger des discussions quant au déplacement d'un carton vers une section particulière du mur de tri, l'animateur vérifie si un ou plusieurs experts sont d'un autre avis. Le cas échéant, une discussion peut avoir lieu pour vérifier les écarts d'opinion. Rappelons qu'avant de déplacer un carton dans la section SÉLECTION, il est essentiel que l'animateur vérifie si le carton en question fait consensus parmi les experts, puisqu'il s'agit d'une décision définitive.

Tout au long de la production collective, l'animateur doit demeurer conscient que chaque participant est constamment en train de confronter ses opinions à celles des autres et de se réajuster en fonction des discussions qui mènent aux déplacements des cartons de section en section, jusqu'à l'atteinte d'un consensus pour une question. Cette confrontation peut être douce et mobilisatrice, mais elle peut également être déconcertante et brutale (Leclerc, 1999). En effet, la production collective est axée sur la sélection des éléments les plus pertinents, impliquant inévitablement que certaines réponses émises seront rejetées. Afin de minimiser les déceptions ou les frustrations chez les participants, l'animateur doit annoncer d'emblée ces situations éventuelles afin de les désamorcer, si possible avec humour. Dans l'introduction, l'animateur devrait mentionner aux experts qu'ils ont émis des réponses de façon individuelle, mais qu'il est tout à fait légitime de se raviser en considérant les réponses des autres. Il est même possible à la lumière des discussions qu'ils en arrivent à considérer sous un autre œil leurs propres réponses et d'en juger certaines moins appropriées.

Le classement des réponses par ordre alphabétique lors de la compilation permet ainsi de conserver la confidentialité de l'auteur, même pour l'animateur. Ce droit à l'anonymat doit être souligné avant d'entreprendre le tri, car il s'avère sécurisant pour les participants. En cas d'ambiguïté à propos d'une réponse d'un participant, l'animateur pourra demander au groupe si la personne qui l'a émise désire expliquer sa pensée. L'auteur de la réponse aura le choix d'apporter certaines précisions et, qu'il le fasse ou non, les décisions seront prises en fonction de l'information disponible.

2.4.3. Terminer la production collective

Le tri d'une question se termine lorsque le nombre d'éléments à retenir dans la section SÉLECTION est atteint, et ce, en fonction du nombre annoncé par l'animateur au début de la production collective. Ceci correspond au consensus du groupe sur la question. À ce moment, les chercheurs prennent note des cartons ou regroupements de cartons présents dans cette section du mur de tri. Pour terminer le tri, il ne doit rester aucun carton dans la section MÉMOIRE COLLECTIVE ni dans la section FRIGO, afin de tenir compte de l'ensemble des opinions émises. Autrement dit, il est essentiel que tous les cartons aient été traités par les experts, que ce soit parce que ces derniers les ont regroupés, scindés ou renommés (REGROUPEMENT), éliminés (POUBELLE) ou, en dernier recours, soumis à d'autres experts en l'absence de consensus (VETO). Un carton peut aussi être transféré seul dans la section REGROUPEMENT, s'il est jugé pertinent par les experts sans qu'il ne puisse être regroupé avec d'autres cartons. Lorsqu'il y a plus d'une question à l'étude, les cartons de la question traitée sont retirés du mur de tri et ceux appartenant à une question subséquente sont affichés dans la section MÉMOIRE COLLECTIVE, afin d'amorcer un nouveau tri selon la même procédure. Selon le nombre de questions à traiter dans la rencontre, cette transition peut s'avérer le moment indiqué pour offrir une pause-santé aux participants. Lorsque toutes les questions ont été triées, la production collective se termine en remerciant les experts de leur temps et énergie consacrés au projet.

2.4.4. Stratégie pour fusionner les résultats issus de plusieurs groupes

Le recours à plusieurs groupes d'experts (cliniciens, gestionnaires, utilisateurs de services, etc.) peut permettre d'explorer un même sujet à l'aide de plusieurs points de vue. En effet, les chercheurs peuvent penser au préalable que les préoccupations sont différentes d'un groupe à l'autre alors que les résultats de TRIAGE montrent le contraire, ou encore confirment ces différences. Par exemple, si l'opinion de cliniciens, de gestionnaires et

d'utilisateurs de services est requise, trois groupes homogènes seront formés, et il n'y aura pas de contact entre les participants des groupes. Trois processus complets de TRIAGE auront lieu de façon distincte, c'est-à-dire que chacune des quatre étapes de TRIAGE décrites précédemment aura lieu avec chaque groupe. Les consensus sont atteints à l'intérieur des groupes, mais on ne vise pas à atteindre un consensus à travers les groupes, puisque les préoccupations peuvent être différentes. Lorsque chacun des trois groupes aura terminé TRIAGE, ce sont les résultats des trois groupes qui seront mis en commun et analysés de façon qualitative par les chercheurs. L'analyse inclut le regroupement des thèmes ou éléments communs à tous les groupes, l'identification des éléments divergents d'un groupe à l'autre et l'explication de ces divergences. Comme il s'agit d'une technique de consensus qui s'appuie directement sur le point de vue des experts de chaque groupe, un retour auprès des participants n'est habituellement pas réalisé après l'analyse des données. En fonction des questions inhérentes à l'étude, les différences ou similarités relevées peuvent conduire à des pistes d'action ou lignes directrices intergroupes ou intragroupes, permettant de répondre aux questions à l'étude en tenant compte de différentes perspectives.

3. EXEMPLES D'UTILISATION DE LA TECHNIQUE

3.1. Adaptation d'un questionnaire autoadministré

Les sections qui suivent présentent une illustration concrète d'une étude ayant eu recours à TRIAGE comme méthode de collecte et d'analyse des données. Cette étude avait pour but d'adapter un questionnaire autoadministré pour qu'il convienne à des travailleurs en incapacité prolongée au travail à la suite de troubles musculosquelettiques (TMS) et admis dans des programmes de réadaptation au travail (Albert, 2012). Une autre étude (Coutu, Durand *et al.*, 2011) avait montré que chez la même population, 50 % des travailleurs absents du travail et admis dans ces programmes présentaient une comorbidité, notamment un trouble mental. Par conséquent, l'illustration retenue dans ce chapitre présente des cas complexes d'incapacité au travail d'origine musculosquelettique, mais pour lesquels un trouble mental a pu se développer.

3.2. Origine de l'étude

Une étude effectuée auprès de cette clientèle avait établi que la représentation, ou autrement dit la façon dont le travailleur comprend son problème de santé, influence grandement sa trajectoire de réadaptation (Baril *et al.*,

2008). Une évaluation systématique de la représentation par les cliniciens permettrait donc d'offrir aux travailleurs des interventions de réadaptation mieux ciblées, favorisant ainsi leur retour au travail (Baril *et al.*, 2008). Parmi les outils de mesure de la représentation recensés, aucun ne convenait à la clientèle cible, mais un questionnaire générique présentait un intérêt pour être adapté (Coutu *et al.*, 2008). Cependant, la littérature scientifique contenait peu d'information sur la nature de l'adaptation, de sorte que l'opinion d'experts devait être sollicitée pour terminer le travail. Par exemple, le questionnaire générique original, intitulé Revised Illness Perception Questionnaire (IPQ-R), évaluait la représentation de la «maladie» selon neuf dimensions. Or les écrits indiquaient que les travailleurs en incapacité prolongée au travail à la suite d'un TMS ne se percevaient pas comme «malades», mais plutôt dans un état intermédiaire, entre la santé et la maladie (Baril *et al.*, 2008; Beaton *et al.*, 2001). Le terme «maladie» dans les items du questionnaire n'apparaissait donc pas approprié pour la clientèle cible, mais un terme plus juste n'avait pas trouvé appui dans la littérature scientifique. De plus, une étude ayant utilisé l'IPQ-R auprès d'une clientèle présentant des TMS recommandait l'adaptation de deux des neuf dimensions du questionnaire, documentant les symptômes ressentis (dimension *Identité de la maladie*) et les causes perçues de la maladie (dimension *Causes*), afin qu'elles reflètent mieux ceux décrits par la clientèle (Foster *et al.*, 2008). En somme, des études suggéraient d'utiliser un questionnaire pour évaluer les représentations, mais aucun outil ne convenait à la clientèle cible. Comme très peu d'écrits étaient disponibles dans la littérature scientifique, l'opinion d'experts devait être sollicitée. En ce sens, TRIAGE a été retenue pour déterminer quelles modifications devaient être apportées au questionnaire existant.

3.3. Préparation

Pour cette étude, les chercheurs se sont d'abord questionnés quant aux experts pouvant fournir l'information recherchée. Le principal critère de sélection était l'expertise clinique ou scientifique dans le domaine de l'incapacité au travail. Le choix s'est arrêté sur six cliniciens et deux chercheurs. Pour les cliniciens, les ergothérapeutes et les psychologues ont été retenus pour le projet puisqu'ils possèdent une bonne connaissance de la sphère psychosociale de la clientèle cible et sont susceptibles d'utiliser le questionnaire dans leur pratique. Pour être sélectionnés comme experts, ils devaient parler couramment français et travailler depuis au moins deux ans dans un milieu où l'objectif d'intervention auprès de la clientèle cible est la réadaptation professionnelle. Les chercheurs devaient quant à eux être affiliés à une université, parler couramment français et avoir effectué des projets de

recherche au cours des trois dernières années sur l'incapacité au travail à la suite de TMS et sur l'élaboration ou la validation d'outils de mesure. Tous les experts recrutés provenaient de la grande région de Montréal, pour des raisons de faisabilité.

Le recrutement s'est fait par téléphone selon une stratégie d'échantillonnage par choix raisonné. Lors du contact, l'intérêt de l'expert à participer à la recherche a été vérifié et le cas échéant, les indications précises concernant l'échéancier lui ont été fournies (dates d'envoi et de retour du formulaire de réponse pour la production individuelle ; date et durée anticipée de la production collective), afin de s'assurer de la disponibilité des experts pour chaque étape. En effet, l'absence d'un expert lors de la production collective peut affecter significativement la procédure de tri, puisque ses réponses issues de la production individuelle sont incluses parmi celles à trier. En l'occurrence, il est alors impossible d'obtenir des clarifications à leur sujet lorsque celles-ci sont nécessaires. Cette situation peut rapidement devenir un irritant pour les participants qui procèdent au tri, si plusieurs réponses de l'expert absent présentent des ambiguïtés qui ne peuvent être résolues.

La deuxième partie de la préparation consistait à préciser les questions à poser aux experts. Pour soutenir leur réflexion, un document préparatoire de cinq pages a été élaboré et contenait : 1) la définition du concept clé de l'étude, soit la représentation de la maladie ; 2) une description de la clientèle cible ; 3) le modèle théorique retenu incluant une définition des neuf dimensions du questionnaire ; et 4) l'outil de mesure original, l'IPQ-R, qui devait être adapté dans le cadre de l'étude (disponible en ligne au : <http://www.uib.no/ipq/pdf/IPQ-R-French.pdf>). Dans cette étude, les experts devaient d'abord se prononcer sur la pertinence du concept de « maladie » et des items formant chacune des neuf dimensions du questionnaire original. Si nécessaire, ils pouvaient ensuite émettre un nombre maximal de suggestions pour remplacer le concept de « maladie » ou perfectionner chacune des dimensions, afin que le questionnaire convienne à une clientèle en incapacité prolongée au travail. La figure 14.2 présente un extrait du formulaire de réponse développé pour l'étude. Pour chaque question, le formulaire contenait une consigne, la question à l'étude, les items sondés suivis d'une échelle de réponse et une section précisant le nombre de suggestions maximal à émettre au besoin (Albert, 2012).

Figure 14.2.
Extrait du formulaire de réponse développé pour l'étude

Consigne : Veuillez maintenant vous référer à la première section de l'IPQ-R, intitulée «**Votre vision de la maladie**» (**dimension** ***Identité***) (p. 1).

Question 2 : LES SYMPTÔMES ÉNUMÉRÉS À LA DIMENSION *IDENTITÉ* SONT-ILS PERTINENTS À SONDER AUPRÈS DE PERSONNES EN INCAPACITÉ PROLONGÉE AU TRAVAIL À LA SUITE DES TMS ?

☑ Veuillez cocher la case qui correspond le mieux à votre réponse.

Symptôme	Pas du tout pertinent	Quelque peu pertinent	Modérément pertinent	Très pertinent	Tout à fait pertinent
1. Douleur	☐	☐	☐	☐	☐
2. Gorge irritée	☐	☐	☐	☐	☐
3. Nausées	☐	☐	☐	☐	☐
4. Difficultés respiratoires	☐	☐	☐	☐	☐
5. Perte de poids	☐	☐	☐	☐	☐
6. Fatigue	☐	☐	☐	☐	☐
7. Articulations raides	☐	☐	☐	☐	☐
8. Yeux irrités	☐	☐	☐	☐	☐
9. Respiration sifflante	☐	☐	☐	☐	☐
10. Maux de tête	☐	☐	☐	☐	☐
11. Lourdeur d'estomac	☐	☐	☐	☐	☐
12. Difficultés pour dormir	☐	☐	☐	☐	☐
13. Vertiges	☐	☐	☐	☐	☐
14. Perte de force	☐	☐	☐	☐	☐

Au besoin, veuillez identifier au maximum dix (10) symptômes qui peuvent être ressentis ou rapportés par des personnes présentant des TMS, ou bien en relation avec ceux-ci. Veuillez inscrire les symptômes suggérés ***en ordre décroissant d'importance***, du plus important (1) au moins important (10).

Symptômes suggérés	
1.	6.
2.	7.
3.	8.
4.	9.
5.	10.

Source : Albert, 2012.

Un prétest de la lettre d'introduction, du document préparatoire et du formulaire de réponse a été effectué auprès de trois cliniciens répondant aux mêmes critères de sélection que les experts. Ils ont reçu exactement la même information que s'il s'agissait de l'envoi officiel aux experts, en ajoutant simplement la consigne suivante: «*Nous vous demandons de faire une lecture très attentive des documents suivants et de répondre aux questions tel qu'indiqué. SVP noter sur les documents toute ambiguïté ou questionnement que vous pourriez avoir tout au long de l'exercice.*» Des modifications mineures ont été apportées aux documents avant de procéder à l'envoi postal marquant le début de la production individuelle.

3.4. Production individuelle

Dans l'envoi postal, une enveloppe préaffranchie sur laquelle était indiquée l'adresse de retour était incluse afin de minimiser la perte de documents. En conformité avec les directives de Dillman *et al.* (2009), un rappel a été envoyé par courriel quelques jours avant la date limite aux experts qui n'avaient pas encore retourné leur formulaire de réponse. Un délai d'environ deux semaines entre la fin de la production individuelle et la rencontre de production collective avait été prévu pour procéder à la compilation des réponses des experts, de sorte que les chercheurs disposaient d'une certaine marge de manœuvre en cas d'envoi tardif de certains experts.

3.5. Compilation

Les formulaires de réponse des huit experts contenaient un total de 153 suggestions. Plus précisément, 26 suggestions ont été émises pour remplacer le concept de «maladie», et 127 suggestions de nouveaux items ont été proposées pour les neuf dimensions du questionnaire. Celles-ci ont été transcrites à l'ordinateur (taille des caractères entre 40 et 72, pour être bien lisibles par les participants), triées par ordre alphabétique, numérotées et imprimées sur des cartons. Une banque de cartons a été générée pour chaque question à l'étude.

3.6. Production collective

La production collective a débuté par l'accueil des participants. La notion de consensus a ensuite été définie comme à la section 1, et la procédure de tri a été expliquée en portant une attention particulière aux situations potentiellement problématiques nommées à la section 2.4.2. Par exemple,

comme l'animateur souhaitait couvrir toutes les questions à l'intérieur de la durée prévue pour la rencontre, il a été expliqué qu'il arrivera à l'occasion qu'il décide de mettre un item qui entraîne des échanges non fertiles au FRIGO, ou encore dans la section VETO si des conflits persistent et que peu de temps est encore disponible. Les cartons d'une première banque avaient été préalablement affichés dans la section MÉMOIRE COLLECTIVE, à gauche du mur de tri. Après avoir demandé aux experts s'ils désiraient obtenir des clarifications concernant le contenu de certains cartons, le tri de la première banque a débuté par la question suivante : « *En vous référant aux cartons de la MÉMOIRE COLLECTIVE, certains peuvent-ils être regroupés ou encore, jugez-vous que certains ne répondent pas à la question de recherche ?* » Les cartons ont ensuite été regroupés, éliminés et sélectionnés par consensus.

Lors de la production collective, les experts ont d'abord procédé au tri des symptômes suggérés pour perfectionner la dimension *Identité* du questionnaire. Les suggestions avaient été émises au bas du formulaire de réponse (voir la figure 14.2). L'équipe de recherche avait préféré débuter par ce tri, puisque le contenu était plus simple et que plusieurs suggestions des experts étaient les mêmes ou très similaires. Plusieurs regroupements seraient donc très faciles à faire, permettant ainsi aux experts d'expérimenter rapidement la procédure de tri. Par exemple, des suggestions telles que « engourdissements », « engourdissements dans les bras », « engourdissements dans les jambes » et « engourdissements/élancements », ou encore « baisse de condition physique », « inactivité forcée », « perte de force/faiblesse », « perte de masse musculaire » et « perte de capacité physique/déconditionnement physique » avaient été émises par différents experts. Ces cartons avaient fait l'objet de deux regroupements, l'un sous le terme « Engourdissements », considéré comme le terme le plus juste par les experts, et l'autre sous l'expression « Diminution de capacité physique ». Ce deuxième regroupement avait donc été renommé suivant la suggestion d'un expert pour que le vocabulaire soit plus accessible aux utilisateurs du questionnaire. Ces deux regroupements avaient ensuite été transférés à la section SÉLECTION, afin d'être conservés dans le questionnaire. Par ailleurs, la suggestion « perte de capacité physique/déconditionnement physique » avait été scindée en deux propositions, puisque les experts considéraient qu'elle contenait deux idées. La suggestion « déconditionnement physique » avait ensuite été regroupée par les experts sous « Manque d'endurance ».

Au total, le tri des 51 suggestions émises pour cette question a nécessité environ 40 minutes, incluant la familiarisation avec la procédure de tri. Pour ce faire, l'animateur a rappelé au groupe à quelques reprises le rôle de chaque section et l'objectif du tri qui était, pour cette étude, de retenir des items qui permettraient de perfectionner un questionnaire autoadministré

s'adressant à des travailleurs en arrêt de travail à la suite de TMS en phase chronique. Les items figurant déjà dans le questionnaire étaient aussi affichés près du mur de tri, à titre de référence.

Lorsque le tri des symptômes a été terminé, les experts s'étaient familiarisés avec la procédure, et le tri des suggestions visant à remplacer le terme de « maladie » dans les items du questionnaire a été amorcé. Au total, 26 suggestions avaient été émises par les experts lors de la production individuelle, dont 18 expressions différentes, comme « accident », « condition de santé », « état actuel » ou « situation de handicap au travail ». Des suggestions comme « physiopathologie persistante » et « syndrome douloureux incapacitant » ont été mis à la POUBELLE par les experts, puisqu'ils étaient formulés dans un langage non accessible pour un questionnaire autoadministré. La suggestion « événement » a aussi été rejetée, car ce terme était considéré comme imprécis. L'objectif du tri était de retenir une seule expression pour remplacer le terme de « maladie » dans les items du questionnaire. Or un long débat a eu lieu concernant le regroupement « Problème de santé » (incluant « difficulté fonctionnelle », « difficulté physique », « problème », « problème musculosquelettique » et « problème physique »), puisque plusieurs experts souhaitaient retenir cette expression, alors que d'autres jugeaient qu'elle avait une connotation négative et qu'elle serait moins pertinente en fin de réadaptation, où l'état de santé du travailleur pourrait s'être significativement amélioré et ne plus représenter un « problème ».

Comme ce débat est survenu très tôt dans le tri, l'animateur avait alors suggéré de mettre ce regroupement au FRIGO, afin d'être en mesure de traiter les autres suggestions figurant dans la section MÉMOIRE COLLECTIVE. Ce faisant, il a été possible de créer un second regroupement avec des suggestions telles que « accident », « blessure », « lésion » et « trouble », mais ce regroupement n'a pas non plus été retenu par les experts, en raison de sa connotation biomédicale et centrée sur le volet physique. Un regroupement a aussi été effectué avec « condition de santé », « état actuel », « état de santé » et « situation ». Ainsi, les suggestions affichées dans la section MÉMOIRE COLLECTIVE ont pu toutes être traitées, soit en étant déplacées dans la section REGROUPEMENT, soit en étant mises à la POUBELLE. À ce moment, l'animateur a réintégré le regroupement « Problème de santé » aux discussions. Une délibération a eu lieu et finalement, l'expression retenue pour remplacer le terme de « maladie » a été « condition de santé actuelle ». Il s'agit de la fusion de suggestions émises par deux experts différents, soit « condition de santé » et « état actuel ». En effet, lors des discussions entre les experts, aucune des suggestions émises ne faisait consensus, de sorte qu'une négociation a eu lieu afin de proposer un nouveau terme, dans ce

cas-ci issu de la fusion de deux suggestions. Il s'agit d'un exemple illustrant comment TRIAGE peut favoriser l'émergence de nouvelles données, grâce aux discussions entre les experts.

Au total, le tri de cette question a nécessité 45 minutes, soit 15 minutes de plus que prévu, en raison des débats qui ont dû avoir lieu pour atteindre un consensus. Or, lorsque les experts ne sont pas d'accord sur un item ou un regroupement, l'animateur doit prendre le temps de questionner les experts sur les raisons sous-jacentes à leur opinion, ce qui permet aux membres du groupe de mieux comprendre les différents points de vue. Cependant, l'enjeu est de permettre aux experts de s'exprimer, tout en évitant de s'enliser dans des débats stériles qui risqueraient d'irriter certains membres et de nuire à la dynamique de groupe. À cet effet, le membre observateur, situé en retrait des discussions, a été très utile pour indiquer discrètement à l'animateur les moments où l'utilisation du FRIGO était nécessaire.

À l'issue de la production collective d'une durée de 3 heures 15 minutes, les experts ont sélectionné par consensus une expression remplaçant le terme de « maladie » parmi les 26 suggestions émises, de même que 26 nouveaux items répartis à travers les 9 dimensions du questionnaire parmi 127 suggestions issues de la production individuelle (Albert, Coutu et Durand, 2013a). Aucun item n'a été placé dans la section VETO. Ce résultat signifie qu'un consensus a véritablement été atteint sur chaque question posée aux experts, puisqu'aucun item n'a soulevé de controverse insoluble. Si cela avait été le cas, il était prévu qu'à la suite de la production collective, deux à trois experts externes au groupe et répondant aux mêmes critères de sélection soient consultés de façon individuelle afin d'obtenir un avis final sur la sélection ou le rejet des items controversés seulement.

À la suite de TRIAGE, le questionnaire adapté a fait l'objet d'un prétest auprès de quelques travailleurs de la clientèle cible et d'une étude de validation, où la cohérence interne et la validité de construit (voir le chapitre 24 de cet ouvrage) ont été documentées (Albert, Coutu et Durand, 2013b). L'outil de mesure adapté, intitulé Questionnaire des Représentations liées à l'Incapacité au Travail (QRIT) est maintenant disponible sur le Web : <http://www.usherbrooke.ca/caprit/fileadmin/sites/caprit/documents/Questionnaire_representation.pdf>.

CONCLUSION

La technique de recherche d'information par l'animation d'un groupe d'experts est une technique encore peu connue en recherche, qui est axée sur l'atteinte de consensus parmi un groupe d'experts. Au même titre que

d'autres techniques de consensus plus répandues comme la technique du groupe nominal ou la méthode Delphi, il ressort que TRIAGE est particulièrement utile lorsque peu d'écrits sont disponibles sur le thème à l'étude. Par contre, TRIAGE se distingue des méthodes plus connues en favorisant les interactions directes entre les participants. L'exemple d'application de TRIAGE a d'ailleurs illustré comment les discussions entre les experts lors de la production collective stimulent à nouveau leur réflexion et permettent l'émergence de nouvelles données. La nécessité pour les experts de traiter tous les cartons avant de terminer la production collective fait aussi en sorte que les consensus tiennent compte de l'ensemble des opinions émises, ce qui est un avantage par rapport aux autres techniques de consensus. En somme, bien que chaque technique de consensus possède ses forces et ses limites, TRIAGE apparaît comme une technique économique et rigoureuse qui gagne à être connue, vu ses multiples usages dans le domaine de la santé mentale et de la réadaptation, lorsqu'une décision éclairée doit être prise parmi un ensemble d'options.

RÉFÉRENCES

Alain, M. et D. Dessureault (dir.) (2009). *Élaborer et évaluer les programmes d'intervention psychosociale*, Québec, Presses de l'Université du Québec.

Albert, V. (2012). *Adaptation et validation d'un questionnaire mesurant les représentations liées à l'incapacité au travail auprès d'adultes présentant un trouble musculosquelettique*, Mémoire de maîtrise, Sherbrooke, Université de Sherbrooke.

Albert, V., M.F. Coutu et M.-J. Durand (2013a). « Adaptation d'un questionnaire visant à mesurer les représentations liées à l'incapacité au travail », *L'Encéphale*, vol. 39, nº 3, p. 174-182.

Albert, V., M.F. Coutu et M.-J. Durand (2013b). « Internal consistency and construct validity of the Revised Illness Perception Questionnaire adapted for work disability following a musculoskeletal disorder », *Disability and Rehabilitation*, vol. 35, nº 7, p. 557-565, <doi:10.3109/09638288.2012.702849>.

Baril, R., M.-J. Durand, M.F. Coutu, D. Côté, G. Cadieux, A. Rouleau et S. Ngomo (2008). *Les représentations de la maladie, de la douleur et de la guérison sur le processus de réadaptation au travail*, Montréal, Institut de recherche Robert-Sauvé en santé et sécurité du travail.

Beaton, D.E., V. Tarasuk, J.N. Katz, J.G. Wright et C. Bombardier (2001). « "Are you better ?" A qualitative study of the meaning of recovery. [Research Support, Non-U.S. Gov't] », *Arthritis Rheum*, vol. 45, nº 3, p. 270-279, <doi : 10.1002/1529-0131(200106)45 :3<270 : :AID-ART260>3.0.CO ;2-T>.

CARRIÈRE, M., S. TÉTREAULT, E.L. BUSSIÈRES et M. GIROUX (2004). *La contribution des gestionnaires des centres de réadaptation en déficience physique à la collaboration entre les intervenants et les parents: une étude multicentrique*, Québec, Université Laval.

COUTU, M.F., M.-J. DURAND, R. BARIL, M.E. LABRECQUE, S. NGOMO et A. ROULEAU (2008). « A review of assessment tools of illness representations: Are these adapted for a work disability prevention context? », *J Occup Rehabil*, vol. 18, p. 347-361.

COUTU, M.F., M.-J. DURAND, A. MARCHAND, M.E. LABRECQUE, D. BERBICHE et G. CADIEUX (2011). *Présence et évolution des facteurs de maintien du trouble de l'anxiété généralisée chez des travailleurs en réadaptation pour une douleur persistante d'origine musculo-squelettique*, Montréal, Institut de recherche Robert-Sauvé en santé et sécurité du travail.

COUTU, M.F., F. LÉGARÉ, M.-J. DURAND, M. CORBIÈRE, D. STACEY, P. LOISEL et L. BAINBRIDGE (2011). « Fostering shared decision making by occupational therapists and workers involved in accidents resulting in persistent musculoskeletal disorders: A study protocol », *Implement Sci*, vol. 6, nº 22, <doi: 10.1186/1748-5908-6-22>.

DELBECQ, A.L., A.H. VAN DE VEN et D.H. GUSTAFSON (1975). *Group Techniques for Program Planning: A Guide to Nominal Group and Delphi Processes*, Glenview, Scott, Foresman.

DILLMAN, D.A., J.D. SMYTH et L.M. CHRISTIAN (2009). *Internet, Mail and Mixed-Mode Surveys: The Tailored Design Method*, 3e éd., Hoboken, John Wiley and Sons.

DRESSLER, L. (2006). *Consensus Through Conversation*, San Francisco, Berrett-Koehler Publishers.

FOSTER, N.E., A. BISHOP, E. THOMAS, C. MAIN, R. HORNE, J. WEINMAN et E. HAY (2008). « Illness perceptions of low back pain patients in primary care: What are they, do they change and are they associated with outcome? », *Pain*, vol. 136, p. 177-187.

GEOFFRION, P. (1997). « Le groupe de discussion », dans B. Gauthier (dir.), *Recherche sociale. De la problématique à la collecte de données*, Québec, Presses de l'Université du Québec, p. 303-328.

GERVAIS, M. (1996). *Étude exploratoire des domaines de référence utilisés par différents acteurs lors de l'évaluation de l'efficacité d'un programme*, Thèse de doctorat, Québec, Université Laval.

GERVAIS, M. et G. PEPIN (2002). « TRIAGE: A new group technique gaining recognition in evaluation », *Evaluation Journal of Australasia*, vol. 2, nº 2, p. 45-49.

GERVAIS, M., G. PEPIN et M. CARRIÈRE (2000). « Triage, un maillage possible entre la recherche et la pratique en ergothérapie », *Revue québécoise d'ergothérapie*, vol. 9, nº 1, p. 11-15.

LECLERC, C. (1999). *Comprendre et construire les groupes*, Québec, Les Presses de l'Université Laval.

PLANTE, J., et M. CÔTÉ (1993). *Technique de recherche d'information par animation d'un groupe expert*. Document non publié.

ST-ARNAUD, Y. (2008). *Les petits groupes : participation et animation*, 3e éd., Montréal, Chenelière Éducation.

SWAINE, B.R., E. DUTIL, L. DEMERS et M. GERVAIS (2003). « Evaluating clients' perceptions of the quality of head injury rehabilitation services : Development and validation of a questionnaire », *Brain Injury*, vol. 17, no 7, p. 575-587.

CHAPITRE 15

L'ANALYSE DE CLASSIFICATION PAR REGROUPEMENT

Description et application à une problématique d'équilibre travail-famille et de détresse psychologique

Alessia Negrini
Jacques Perron
Bertrand Perron[1]

FORCES

- Elle facilite l'exploration d'observations de tous genres en vue de les classer en regroupements distincts.
- Elle permet de constituer des typologies en ciblant et décrivant des groupes d'individus définis par des similarités sur de multiples dimensions d'intérêt pour un domaine d'application.
- Elle offre plusieurs algorithmes de classification qui se prêtent à diverses configurations de données.

LIMITES

- Elle n'est pas fondée sur une théorie unique généralement reconnue.
- Elle est sensible à plusieurs enjeux empiriques tels que la présence de cas marginaux, le nombre et la taille des regroupements et le nombre de variables à introduire dans l'analyse.
- Elle laisse une marge importante au chercheur pour l'analyse des résultats, ce qui peut mener à des interprétations biaisées quant aux conclusions théoriques.

---

1. Les trois auteurs remercient la commission scolaire du Chemin-du-Roy (Trois-Rivières, Québec) pour la contribution apportée à cette production scientifique.

Ce chapitre porte sur la méthode d'analyse de classification par regroupement. Il est d'abord consacré à la description de cette méthode statistique, plus précisément dans ses variantes les plus fréquemment utilisées en psychologie pour la classification des personnes. Puis, une démarche complète d'analyse est présentée dans le but de montrer un exemple d'application de cette méthode. À partir de données recueillies auprès du personnel d'une organisation publique québécoise, des groupes distincts de personnes en termes d'équilibre travail-famille ont été constitués et, par la suite, comparés selon leur niveau de détresse psychologique.

1. ORIGINES, OBJECTIF ET VARIANTES DE LA MÉTHODE

En 1939, Tryon effectue le premier travail systématique sur l'analyse par regroupement. Ses travaux subséquents sont publiés à partir des années 1950 (Tryon, 1955), et de 1960 à 1970 (Tryon et Bailey, 1970), la méthode commence à attirer l'attention des chercheurs (Blashfield et Aldenderfer, 1988) probablement à cause de la vitesse d'exécution accrue des ordinateurs. Par ailleurs, la publication par deux biologistes (Sokal et Sneath, 1963) de l'ouvrage *Principles of Numerical Taxonomy* fut aussi déterminante, en mettant en évidence la classification d'organismes vivants similaires en regroupements homogènes.

Utilisée dans de nombreuses disciplines scientifiques, l'analyse par regroupement se caractérise par plusieurs algorithmes de classification différents (Scoltock, 1982). Malgré leurs spécificités, ils ont tous comme objectif de faire ressortir la structure de groupes d'une banque de données (Milligan et Hirtle, 2003) en établissant que les membres d'un même regroupement se ressemblent entre eux tout en étant différents des membres des autres regroupements (Henry, Tolan et Gorman-Smith, 2005). Cette méthode peut donc être appliquée à des situations différentes qui impliquent une activité de classification. À titre d'exemple, en psychologie du travail, elle permet d'établir des profils de travailleurs (p. ex., ayant un bon état de santé psychologique ou état précaire), de définir des modèles de comportements (p. ex., présence au travail ou absence) et de cibler les interventions sur des éléments précis (p. ex., prévenir l'absence du travail de personnes ayant un état précaire de santé psychologique).

Les méthodes d'analyse de classification par regroupement peuvent être divisées en quatre catégories sur la base de l'algorithme utilisé : hiérarchique, non hiérarchique, chevauchante (algorithme basé sur la densité) et algorithme de grille (Halkidi, Batistakis et Vazirgiannis, 2001). Le choix de l'algorithme à utiliser dépend, entre autres, de son efficacité de traitement informatique des données, de la qualité des regroupements obtenus, de la

simplicité de l'interprétation des résultats et du type d'échelle de mesure des variables (Saitta, Raphael et Smith, 2008). Dans ce chapitre, la méthode hiérarchique et la méthode (non hiérarchique) par nuées dynamiques seront illustrées.

1.1. Étapes de la méthode d'analyse de classification par regroupement

Plusieurs auteurs proposent des séquences similaires d'étapes à suivre pour réaliser l'analyse de classification par regroupement (Henry *et al.*, 2005; Lorr, 1983; Milligan et Hirtle, 2003; Rapkin et Luke, 1993). La séquence suggérée ici est une adaptation de ces différents travaux pour distinguer les étapes selon le moment de leur application: 1) préparation; 2) exécution; 3) description, validation et interprétation des résultats.

1.2. Préparation

1.2.1. Choix des observations

En fonction des buts poursuivis sont d'abord choisies les caractéristiques des personnes faisant l'objet de l'analyse (p. ex., nombre, sexe, âge, type d'emploi) et, ensuite, les variables utilisées pour la classification. Milligan et Hirtle (2003) soulignent que cette étape est délicate, car seules les variables qui contribuent à distinguer des groupes devraient être incluses dans l'analyse. Sachant qu'un algorithme de regroupement peut toujours générer une structure de groupes à partir d'une banque de données (Henry *et al.*, 2005), il importe de montrer que le choix des variables est théoriquement pertinent au regard du domaine d'application. Il faut aussi réaliser que l'ajout d'une ou deux variables non pertinentes peut modifier de façon significative la structure de groupes émanant des données. S'il n'y a pas de consensus sur la participation d'une variable au construit théorique d'application, on peut effectuer l'analyse de regroupement, d'abord avec et ensuite sans cette variable, puis comparer les deux solutions.

Bien qu'il soit possible de faire des analyses de classification par regroupement à partir de variables catégorielles, des variables continues ou ordinales sont habituellement utilisées dans le domaine des sciences sociales pour mesurer, par exemple, des perceptions et des comportements des individus. Il n'est pas nécessaire que les variables choisies se distribuent normalement ou qu'elles aient une faible corrélation entre elles, car plusieurs algorithmes s'accommodent de telles caractéristiques pour établir la structure de groupes sous-jacente à une banque de données (Milligan et Hirtle, 2003).

1.2.2. Homogénéisation des mesures

Dans une analyse par regroupement, les variables dont l'étendue et la variance sont plus élevées ont une influence accrue sur la formation des groupes (Henry *et al.*, 2005). S'il n'est pas possible de justifier théoriquement les pondérations dues à la variation des étendues, il est alors recommandé de normaliser les variables introduites dans l'analyse. Pour que leur poids soit égal lors des calculs à venir, on peut transformer les scores en les ramenant à une même échelle de mesure commune à toutes les variables (p. ex., de 1 à 5).

1.2.3. Examen des données

L'examen des données peut aider à orienter le choix de la méthode de regroupement tout comme il pourra s'avérer utile à l'interprétation de la solution finale. La solution de regroupement peut être influencée par certaines variables présentant des distributions particulières. C'est le cas, par exemple, d'une distribution de forme exponentielle dont la catégorie modale se situe à une extrémité de l'échelle de mesure, comparativement à une distribution normale en forme de cloche. L'analyse des distributions est donc recommandée, d'autant plus qu'elle peut servir à repérer des réponses extrêmes et, s'il y a lieu, à éliminer un ou plusieurs participants dont les réponses sont extrêmes ou atypiques. Aussi, l'envergure des corrélations entre les variables permet de vérifier leur indépendance relative : des coefficients élevés entre certaines variables accentuent leur poids dans les calculs à venir (Rapkin et Luke, 1993).

1.2.4. Choix de la méthode d'analyse par regroupement

Que la méthode soit hiérarchique ou non hiérarchique, le calcul des distances entres les individus sert à regrouper ceux qui sont les plus similaires en classes distinctes. Toutefois, ce processus général ne se concrétise pas dans les mêmes conditions selon la méthode de regroupement choisie.

La *méthode hiérarchique* propose une démarche de nature inductive au sens où les analyses sont réalisées sans avoir à formuler *a priori* une hypothèse quant à la constitution des regroupements ; ceux-ci sont définis au fur et à mesure que sont agrégées les comparaisons de distances entre toutes les paires d'individus. Ainsi, pour un échantillon de n individus, selon la formule $n*(n - 1)/2$ paires distinctes, le nombre de comparaisons peut devenir élevé si n est grand (proportionnel au carré de n). La taille de

l'échantillon entre donc définitivement en ligne de compte dans le choix de cette méthode. Exécutée sur une banque de données de grande taille (1000+), elle peut occasionner des temps de traitement informatique excessifs.

La méthode hiérarchique agglomère les individus par regroupements successifs et mutuellement exclusifs jusqu'à la formation d'un seul groupe qui les englobe tous (algorithme par agglomération). Elle peut aussi procéder par divisions successives, partant d'un seul groupe jusqu'à la restitution de tous les individus en groupes singuliers (algorithme par division) (Jain, Murty et Flynn, 2000). La structure toujours nichée des solutions suggérées par la méthode hiérarchique se lit sur un dendrogramme, c'est-à-dire une représentation graphique par laquelle il est possible d'observer les diverses possibilités de regroupements.

La *méthode par nuées dynamiques* fonctionne de manière plutôt hypothético-déductive. Partant de centres de classes initiaux d'un nombre préétabli de regroupements, elle cherche, en apportant des ajustements par itérations successives, à atteindre un point de convergence correspondant objectivement à l'arrêt de changements dans l'ajustement des centres. Chaque fois qu'un individu est attribué à un groupe, le centre ou centroïde de ce groupe est immédiatement recalculé. Comparativement aux méthodes hiérarchiques, la convergence est assurée de façon plus rapide, avec un plus petit nombre d'itérations. Ces opérations répétées permettent de déterminer les centres de classes finaux et le nombre de personnes dans chacun des regroupements (Jain *et al.*, 2000).

Enfin, lorsque la taille de l'échantillon le permet, l'usage combiné des deux méthodes est un choix intéressant en ce qu'il permet de comparer deux solutions en prenant les centres obtenus avec la méthode hiérarchique comme point de départ de l'application de la méthode par nuées dynamiques. Cette méthode mixte est souvent utilisée en psychologie (p. ex., Demerouti et Geurts, 2004; Henry *et al.*, 2005).

1.2.5. Choix du logiciel

Quelques logiciels (p. ex., SAS, STATA, SPAD, SPSS) permettent d'exécuter les analyses de classification par regroupement (Everitt, Landau et Leese, 2001). Du fait qu'il en existe une version française, qu'il est particulièrement adapté aux besoins d'analyses en sciences sociales et de la santé, qu'on le retrouve dans plusieurs établissements d'enseignement et de recherche, et qu'il offre trois options d'analyses de classification, le logiciel SPSS (Statistical Package for the Social Sciences [IBM Corporation, 2012]) est donc un choix approprié. Comme cela est illustré au tableau 15.1, plusieurs éléments sont à considérer dans le choix de la méthode d'analyse par regroupement.

Tableau 15.1.
Éléments à considérer dans le choix de la méthode d'analyse par regroupement

Éléments	**Méthode**		**Commentaires**
	Hiérarchique	**Par nuées dynamiques**	
Taille de l'échantillon	Temps d'exécution excessif si l'échantillon est de grande taille.	Peut être exécutée sur des échantillons de grande taille.	En méthode hiérarchique, il est concevable de sélectionner aléatoirement des échantillons de cas d'une très grande banque de données et de mener des analyses dont la comparaison facilite la décision quant à la solution finale à retenir.
Nombre de groupes	Spécification préalable non requise. Il est tout de même possible de forcer l'arrêt des agglomérations ou divisions en spécifiant le nombre de groupes voulu.	Spécification préalable requise.	
Mesure de distance	À choisir parmi des possibilités offertes par SPSS.	Distance euclidienne simple préintégrée dans la procédure SPSS.	La mesure euclidienne est la plus souvent utilisée. Recommandée si les variables introduites dans l'analyse procèdent toutes d'une même échelle (Henry *et al.*, 2005).
Cas extrêmes	En mode par agglomération, les cas extrêmes demeurent longtemps des groupes singuliers avant de s'assimiler à un groupe.	L'analyse peut être sensible à la présence de cas extrêmes.	Il est recommandé, si théoriquement justifié, de les retirer. Leur présence peut distordre la solution de la méthode par nuées dynamiques, surtout si le choix aléatoire des centres de regroupements initiaux concorde avec un cas extrême. Des centres de regroupements prédéterminés permettent d'éviter cet écueil.
Initialisation des centres de classes	Ne procède pas initialement par mesure de distance par rapport à des centres de regroupements.	Si pertinent, prédéterminer les centres de regroupements. SPSS n'offre pas cette option. Alors, il faut modifier la syntaxe pour inscrire les centres de classes initiaux.	Les centres de regroupements peuvent être prédéterminés dans la méthode non hiérarchique à partir de types-idéaux suggérés par les connaissances du domaine d'application. Ils correspondent aux moyennes de chacune des variables de chaque regroupement.

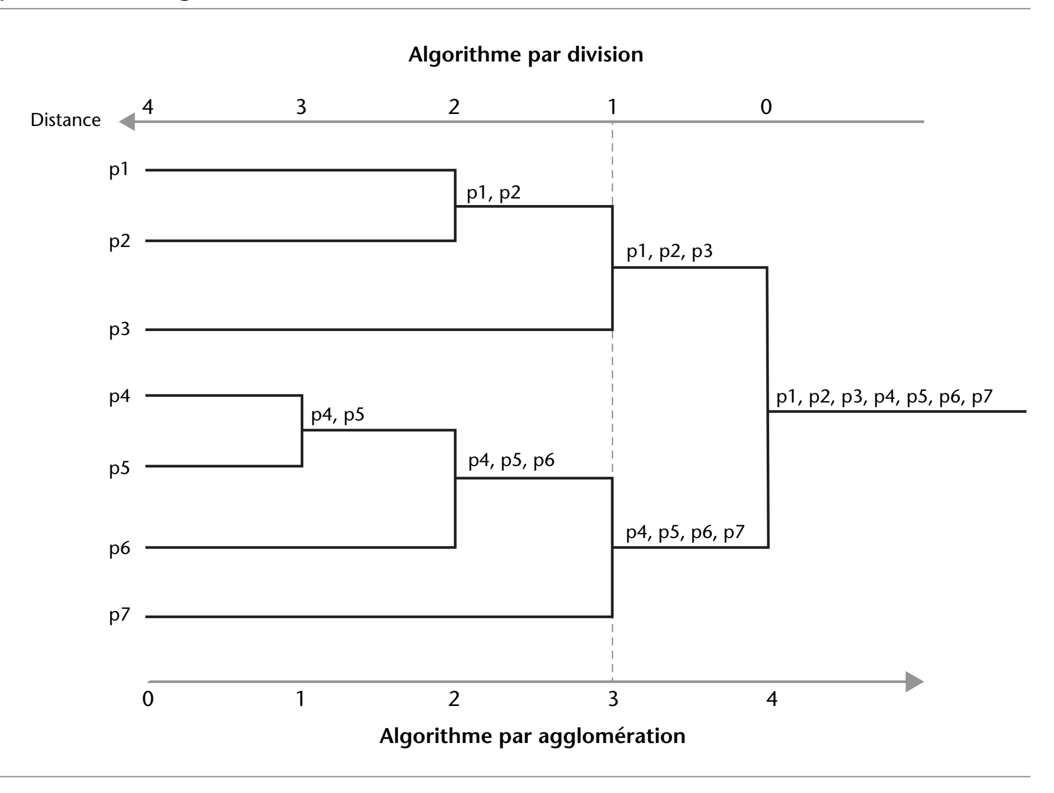
1.3. Exécution

1.3.1. Méthode hiérarchique

Dans le volet *Analyse* de la version 21 de SPSS (IBM Corporation, 2012), le répertoire *Classification* comprend la *Classification hiérarchique*, qui comporte à son tour cinq sections : *Variables*, *Statistiques*, *Diagrammes*, *Méthode* et *Enregistrer*. Les choix suivants permettent toutefois d'exécuter les étapes essentielles de la méthode. Dans l'espace prévu, inscrire toutes les *Variables* servant à constituer les regroupements. Parmi les options du volet *Statistiques*, sélectionner *Chaîne des agrégations* et *Matrice des distances*. Par la suite, en cliquant sur *Diagrammes*, il est possible de choisir le format de la représentation graphique des résultats (c.-à-d. *Arbre hiérarchique*). Dans la section *Méthode*, choisir la *Méthode d'agrégation de Ward*, ainsi que la *Mesure de la distance euclidienne au carré*. En analyse hiérarchique, la méthode de Ward (1963) est reconnue pour son adaptation à de multiples formes de groupes. La distance euclidienne est utilisée pour les données mesurées sur des échelles à intervalle. Le carré de la distance euclidienne assure que les distances les plus grandes ont le plus d'importance. L'exécution de ces commandes affiche un ensemble de résultats, dont le principal est un dendrogramme (*Arbre hiérarchique*) qui sert à voir tous les regroupements.

La figure 15.1 illustre les résultats fictifs de classification hiérarchique de 7 personnes (p1 à p7) à partir de leurs réponses à *n* variables. Chaque branche du dendrogramme (ligne verticale) correspond à un regroupement. La ligne horizontale qui relie deux ou plusieurs branches indique le niveau de la distance à laquelle les regroupements se fusionnent ou se divisent. En effet, le dendrogramme peut être lu de deux façons : des branches vers le tronc (→algorithme par agglomération) ou du tronc vers les branches (←algorithme par division). Sur le graphique constitué, on observe les différents niveaux de coupe de la hiérarchie pour dégager les solutions possibles et le nombre de groupes qui les composent. Pour ce faire, on sectionne l'arbre à proximité d'un embranchement (ligne verticale pointillée). Si on interprète le dendrogramme ci-dessus selon la séquence par agglomération et la ligne de coupure (pointillée) suggérée, la solution la plus plausible en est une à deux regroupements, formés chacun de trois (p1, p2, p3) et de quatre (p4, p5, p6, p7) personnes, à la fois homogènes sur le plan interne et nettement distincts l'un de l'autre. À noter que ce dendrogramme ne suggère pas de solution à trois regroupements puisque le précédent point de coupure applicable suggère quatre groupes, dont deux sont constitués d'individus seuls (p7 et p3).

Figure 15.1.
Exemple de dendrogramme



Après avoir dégagé du dendrogramme la ou les solutions plausibles, il est de mise de reprendre l'analyse avec SPSS et d'*Enregistrer la ou les Classes d'affectation* pour les solutions retenues, ce qui permettra la description des regroupements dans des analyses ultérieures.

1.3.2. Méthode par nuées dynamiques

Dans le volet *Analyse* de SPSS, le répertoire *Classification* comprend la *Classification en nuées dynamiques*, qui comporte quatre sections : *Variables*, *Itérer*, *Enregistrer* et *Options*. Pour commencer, dans l'espace prévu, inscrire, en valeurs brutes, toutes les *Variables* servant à constituer les regroupements ; inscrire un *Nombre de classes* (regroupements) à constituer ; choisir la méthode à suivre dans les options *Itérer et classer* et les *Centres de classes* à lire.

Itérer permet de sélectionner le *Maximum des itérations* au choix (par défaut = 10) ainsi que le *Critère de convergence* (par défaut = 0). *Enregistrer la ou les Classes d'affectation*. Dans le volet *Options*, sélectionner les statistiques suivantes : *Centres de classes initiaux*, *Tableau ANOVA*, *Affectation et distance aux centres*. En fonction de ces commandes, dans deux nouvelles colonnes de la banque de données sont ajoutées pour chaque personne l'affectation à son regroupement d'appartenance et sa distance au centre de ce regroupement.

Les informations suivantes sont affichées dans la section des résultats : *a*) à chacun des regroupements, les centres de classes initiaux et finaux pour chacune des variables inscrites ; *b*) un historique des itérations ; *c*) pour chaque membre de l'échantillon, son regroupement d'appartenance et sa distance au centre de ce regroupement ; *d*) les résultats d'analyses de variance aux variables inscrites en fonction des regroupements constitués.

1.4. Description, validation et interprétation des résultats

Quelle que soit la méthode choisie, il est important de décrire les groupes constitués, de démontrer la validité de la solution retenue et de proposer une interprétation des groupes théoriquement fondée.

1.4.1. Description des résultats

Pour décrire les regroupements obtenus, il faut d'abord s'assurer qu'un code d'appartenance à l'un ou l'autre de ceux-ci est attribué à chacun des individus présents dans la banque de données. En complément à l'analyse de classification par regroupement, il devient alors possible d'effectuer une analyse de variance avec comparaisons de moyennes sur les variables constitutives des groupes. Étant donné que l'analyse par regroupement cherche toujours à maximiser l'homogénéité interne des groupes et l'hétérogénéité entre les groupes, l'analyse de variance croisant la typologie et les variables la constituant donnera toujours des résultats valides sur le plan statistique (tests F et test d'homogénéité). Il s'agit donc d'une méthode servant davantage à décrire qu'à valider les résultats (Aldenderfer et Blashfield, 1984 ; Khalid et Bouklouze, 2006).

1.4.2. Validation des résultats

Il existe de nombreuses façons d'examiner la validité interne et externe des résultats d'une analyse de classification par regroupement.

Validité interne

Des analyses complémentaires sont de nature à apporter des informations additionnelles quant à la validité d'une solution d'analyse de classification par regroupement. Une autre façon de corroborer une solution consiste à scinder aléatoirement l'échantillon initial et à procéder à des analyses répétées à l'aide de l'une ou l'autre des deux méthodes de classification par regroupement (Luke, Rappaport et Seidman, 1991).

Aussi, lorsque deux méthodes de classification en arrivent à des regroupements conceptuellement semblables, mais différents quant à leurs effectifs, il est indiqué d'approfondir davantage le choix de l'une ou de l'autre, surtout en cas d'inférence à une population générale. Ainsi, le recours à des indices statistiques permettant de juger de la qualité des solutions obtenues peut aider au choix de la solution finale à retenir (Hennig, 2008; Milligan, 1981; Rujasiri et Chomtee, 2009; Schwarz, 1978).

Validité externe

La différenciation des regroupements selon des variables non impliquées dans la définition de la typologie constitue une forme de validation externe de la solution retenue (Moran *et al.*, 2012). Encore une fois, les résultats du croisement avec les variables externes doivent être défendus théoriquement pour prétendre avoir validé la solution. Ainsi, le fait de montrer que des regroupements définis sur la base de leurs caractéristiques (p. ex., divers niveaux d'équilibre travail-famille) diffèrent l'un de l'autre en termes de santé psychologique (p. ex., détresse psychologique) constitue un exemple de ce type de validation. D'autres variables externes (p. ex., sexe, âge, type d'emploi) dans une organisation peuvent aussi servir à distinguer les regroupements les uns des autres.

1.4.3. Interprétation des résultats

À un premier niveau, il s'agit, compte tenu du profil de résultats obtenus à l'ensemble des variables, de conférer une signification et d'assigner un vocable ou une expression à chacun des regroupements de sorte qu'ils participent d'un même construit (p. ex., équilibre travail-famille). La normalisation de la mesure des variables et la représentation graphique des résultats facilitent la visualisation des données à interpréter. À un second niveau, les résultats obtenus (p. ex., nombre de regroupements, configuration des variables constituant le profil de chaque regroupement) sont mis en lien avec une théorie ou avec d'autres résultats déjà connus en rapport avec la problématique étudiée.

2. PRÉSENTATION D'UNE APPLICATION DE L'ANALYSE DE CLASSIFICATION PAR REGROUPEMENT

2.1. Préparation

Comme décrit à la section 1.2.1, les auteurs ont effectué plusieurs choix (cadre conceptuel, échantillon, variables) pour bien répondre aux objectifs de l'étude.

2.1.1. Cadre conceptuel : équilibre travail-famille

La notion d'équilibre travail-famille (pour une synthèse : Greenhaus et Allen, 2010) sert à décrire les relations entre ces deux domaines de vie qui, au cours des trente dernières années, ont été analysés en termes d'interférence, de conflit et de facilitation. Selon Carlson et Grzywacz (2008), l'interférence, c'est-à-dire l'empiètement d'un rôle sur l'autre, est le phénomène qui a été le plus fréquemment étudié. Par ailleurs, le conflit renvoie spécifiquement à une incompatibilité de pressions provenant des deux rôles. Enfin, objet de travaux plus récents, la facilitation correspond à l'idée que le travail et la famille puissent être réciproquement bénéfiques. À cause de l'instrumentation utilisée et de l'approche conceptuelle sur laquelle elle s'appuie, la notion d'interférence entre le travail et la famille est celle qui est retenue pour ce chapitre.

2.1.2. Choix des variables

À la suite des travaux de Frone, Russell et Cooper (1992), le consensus qui rallie le plus de chercheurs veut qu'entre les rôles des deux domaines de vie l'interférence est bidirectionnelle, que les deux types d'interférence sont corrélés et qu'ils doivent être pris en compte simultanément. Les variables *Travail interférant avec famille* (TIF) et *Famille interférant avec travail* (FIT) sont donc les premiers éléments retenus ici pour constituer une typologie. Dorio, Bryant et Allen (2008) font état de trois méta-analyses qui concluent à un ensemble de relations négatives de TIF avec la *Satisfaction au travail* (ST). Des relations négatives ont aussi été observées entre FIT et la *Satisfaction familiale* (SF) (Kinnunen, Geurts et Mauno, 2004 ; Parasuraman *et al.*, 1996). Les variables ST et SF font aussi partie de la classification par regroupement dans ce chapitre. Citant entre autres la contribution de Rantanen, Pulkkinen et Kinnunen (2005), Mullen, Kelley et Kelloway (2008) soulignent que la *Détresse psychologique* est significativement associée aux deux formes

d'interférence du travail avec la famille (c.-à-d. TIF et FIT). C'est pourquoi la *Détresse psychologique* sera analysée à titre de variable externe qui caractérise les regroupements de la typologie proposée.

2.1.3. Objectifs

Dans le prolongement des travaux de Demerouti et Geurts (2004), il s'agit de répondre à la question suivante: «L'équilibre entre le travail et la famille est-il vécu différemment par divers groupes de personnes dont la santé psychologique se situe aussi à des niveaux variables?» Par conséquent, en réponse à cette question, nous escomptons dégager des regroupements conceptuellement interprétables et constitués par combinaisons variables de niveaux d'interférence (TIF et FIT) et de satisfaction (ST et SF). À partir des regroupements d'une telle typologie, on s'attend aussi à observer des différences significatives de moyennes montrant que la *Détresse psychologique* augmente en fonction des niveaux d'envergure de TIF ou de FIT ou des deux combinés à de faibles niveaux de ST ou de SF ou des deux.

2.1.4. Échantillon

Dans le cadre d'une plus vaste étude, les participants ont anonymement répondu par courrier à un ensemble de questionnaires relatifs au bien-être et à la santé psychologique (Perron, Lamothe et Daoust, 2005). Selon le sexe, la fonction et l'âge, l'échantillon aléatoire de 389 participants représente dans des proportions similaires l'ensemble des 2500 membres du personnel d'une commission scolaire du Québec. Il se compose de 252 femmes (64,8%) et de 137 hommes (35,2%). Variant de 23 à 75 ans, l'âge moyen de cet échantillon est de 42,81 ans (é.t. = 9,40).

2.1.5. Échelles de mesure

Les mesures d'interférence entre les deux domaines de vie (travail et famille) les plus souvent utilisées sont celles de Kopelman, Greenhaus et Connolly (1983), reprises par Parasuraman *et al.* (1996). Elles ont été traduites et adaptées par Perron (1999) à l'aide d'échantillons québécois. Pour chaque item, l'évaluation se fait à l'aide d'une échelle de Likert en cinq points variant de 1 = pas du tout d'accord à 5 = tout à fait d'accord. La variable *Travail interférant avec famille* (TIF) a été mesurée à l'aide de six items et celle *Famille interférant avec travail* (FIT) à partir de quatre items.

La mesure de *Satisfaction familiale* (SF) (Kopelman *et al.*, 1983) a été définie par trois items. Pour mesurer la *Satisfaction au travail* (ST), la version française (Fournier, 1995) du Minnesota Satisfaction Questionnaire (MSQ; Weiss *et al.*, 1967) a été utilisée. Il s'agit d'un ensemble de 20 items illustrant autant de facettes caractéristiques d'un emploi et évalués à l'aide d'une échelle en cinq points allant de 1 = très insatisfait(e) à 5 = très satisfait(e).

L'Indice de détresse psychologique de Santé Québec (IDPESQ14) (Légaré *et al.*, 2000) comprend 14 items. En référence à la *dernière semaine*, les réponses aux questions se font à l'aide d'une échelle variant de 0 = jamais à 3 = très souvent.

Les coefficients de cohérence interne de ces cinq échelles sont satisfaisants (tableau 15.2; alpha de Cronbach > 0,70). Ainsi, puisqu'ils dénotent la fiabilité des échelles et entrent dans le calcul de l'erreur de mesure, ils contribuent à la reproductibilité des résultats de l'analyse de classification par regroupement.

2.1.6. Examen des données

Des analyses statistiques préliminaires sur les données ont été effectuées (p. ex., homogénéisation des mesures, analyse des distributions) pour décrire l'échantillon et pour orienter le choix de la méthode de regroupement à utiliser.

Tableau 15.2.
Statistiques descriptives, matrice de corrélations et cohérence interne des variables à l'étude

Variables	Moyenne	Écart-type	1.	2.	3.	4.	5.
1. TIF	2,91	0,93	*0,86*				
2. FIT	1,70	0,57	0,27	*0,74*			
3. ST	3,85	0,53	–0,33	–0,16	*0,88*		
4. SF	4,39	0,77	–0,23	–0,33	0,17	*0,86*	
5. Détresse	0,68	0,46	0,43	0,27	–0,32	–0,33	*0,89*

Notes : $n = 389$; pour toutes les corrélations, $p < 0,01$.
Les coefficients de cohérence interne (alpha de Cronbach) des échelles sont représentés sur la diagonale, en caractères italiques.

Les données du tableau 15.2, prélevées auprès de 389 participants, portent sur les variables choisies pour constituer les regroupements et en étudier la validité externe. Comme dans les travaux appuyant le choix des variables, il est habituel de remarquer des moyennes moins élevées aux variables d'interférence (TIF et FIT) qu'à celles de satisfaction (ST et SF). Par ailleurs, la relative indépendance de ces variables se constate au faible niveau de leurs intercorrélations (de –0,16 à –0,33) et laisse entrevoir qu'elles auront un pouvoir spécifique de différenciation des regroupements.

2.1.7. Choix du logiciel et de la méthode de classification

SPSS version 21 (IBM Corporation, 2012) est le logiciel utilisé pour les analyses présentées dans cette partie du chapitre. Toutes les commandes disponibles dans SPSS ne seront pas présentées de façon détaillée. Seulement les étapes d'exécution qui ont présidé aux choix des résultats seront illustrées.

En ce qui concerne la méthode de classification, un choix mixte est préconisé puisque la méthode hiérarchique est appliquée en alternance avec la méthode par nuées dynamiques.

2.2. Exécution

Plusieurs analyses par regroupement ont été exécutées aux fins du présent exercice. D'abord, une analyse hiérarchique a été effectuée afin de dégager les solutions de regroupements possibles. Les choix faits dans le cadre de l'analyse hiérarchique (*Méthode d'agrégation de Ward*, *Mesure de la distance euclidienne au carré*, *génération du Dendrogramme*) ont été présentés à la section 1.3. Ensuite, sur la base des résultats obtenus de l'analyse hiérarchique, des analyses de classification par nuées dynamiques ont été effectuées en laissant le logiciel SPSS établir les centres de classes initiaux de regroupement de façon aléatoire. Aussi, pour appuyer le choix de la solution finale, les résultats de ces deux types d'analyses ont été comparés. Nous avons dû reprendre l'analyse de classification hiérarchique en commandant cette fois-ci à SPSS d'attribuer les classes d'affectation pour la solution semblant la plus plausible. Enfin, une dernière analyse par nuées dynamiques a été effectuée en initiant comme centres de classes ceux obtenus de l'analyse hiérarchique précédente (Andes, 1998, p. 39; Steinley, 2006, p. 7). En dernier lieu, pour valider la solution obtenue, les analyses ont été reprises sur deux sous-échantillons générés au hasard à partir de l'échantillon initial.

2.3. Description, validation et interprétation des résultats

2.3.1. Description des résultats

Pour la description des résultats, les informations concernant l'appartenance de chaque membre de l'échantillon à son regroupement et la distance au centre de ce regroupement consistent en des tableaux trop imposants pour être reproduits dans ce chapitre. Les auteurs les transmettront à la demande aux lecteurs intéressés. La description des variables est présentée au tableau 15.2 (scores bruts) ainsi qu'au tableau 15.7 et à la figure 15.2 (scores normalisés).

2.3.2. Classification hiérarchique

Dans SPSS, les résultats se présentent sous forme d'indices statistiques et de représentations graphiques. Dans notre cas, l'analyse du dendrogramme généré a montré une structure comptant de deux à cinq regroupements. L'étape suivante consiste à utiliser ces résultats comme base pour spécifier le nombre de regroupements à considérer dans l'analyse par nuées dynamiques.

2.3.3. Classification par nuées dynamiques

Des analyses par nuées dynamiques ont été effectuées sur la base des solutions suggérées par l'analyse hiérarchique précédente en spécifiant alternativement de deux à cinq groupes. Les résultats de ces solutions ont été comparés. La solution à trois groupes a semblé être la plus plausible tant sur le plan empirique (nombre et taille des groupes) que sur le plan théorique (profils au regard de la littérature). Quant aux autres solutions, dans celle à deux regroupements, seulement la variable TIF contribue fortement à distinguer les regroupements. Pour les solutions à quatre et à cinq regroupements, les résultats d'analyse de variance (tests F) montrent, entre les regroupements, des distinctions moins nettes que dans le cas de la solution à trois regroupements. C'est particulièrement le cas en ce qui concerne les deux variables (TIF, SF) qui contribuent le plus à la formation des regroupements. De plus, les solutions à quatre et à cinq regroupements comprennent un regroupement aberrant sur le plan théorique, du fait que tous ses centres de classes se trouvent au-dessus de la moyenne générale de chacune des quatre variables constituantes.

Les résultats présentés ci-dessous sont donc ceux de la méthode par nuées dynamiques à trois regroupements, sans spécification préalable des centres de classes (avec initialisation aléatoire). À partir du compte rendu de SPSS, on trouve les centres de classes initiaux déterminés aléatoirement (tableau 15.3). Les données du tableau 15.4 indiquent que les changements dans les centres de classes ont atteint le critère de convergence à la huitième itération. Au tableau 15.5 se trouvent les centres de classes finaux et le nombre de personnes dans chaque regroupement.

Tableau 15.3.
Analyse par nuées dynamiques avec initialisation aléatoire : centres de classes initiaux

	Classe		
	1	2	3
TIF	1,00	4,83	3,67
FIT	1,00	1,25	3,50
ST	4,40	3,50	3,65
SF	5,00	5,00	2,00

Tableau 15.4.
Analyse par nuées dynamiques avec initialisation aléatoire : historique des itérations

	Changements dans les centres de classes		
Itération	**1**	**2**	**3**
1	1,257	1,365	1,537
2	0,027	0,055	0,161
3	0,000	0,041	0,092
4	0,000	0,015	0,033
5	0,000	0,022	0,048
6	0,019	0,027	0,026
7	0,020	0,019	0,000
8	0,000	0,000	0,000

Tableau 15.5.
Analyse par nuées dynamiques avec initialisation aléatoire : centres de classes finaux

	Classe		
	1 (*n* = 159)	2 (*n* = 160)	3 (*n* = 70)
TIF	2,01	3,57	3,44
FIT	1,49	1,75	2,09
ST	4,06	3,72	3,66
SF	4,69	4,69	3,03

Avant d'en venir au choix final de la solution à trois regroupements, d'autres analyses ont été réalisées (voir section 2.3.4.).

2.3.4. Comparaison des résultats : méthode hiérarchique et méthode par nuées dynamiques

Après avoir statué, à partir des analyses par nuées dynamiques, que la solution à trois groupes semblait la plus appropriée, deux comparaisons ont été effectuées. D'abord, au tableau 15.6, on voit que les résultats obtenus par la méthode hiérarchique et la méthode par nuées dynamiques (avec centres de classes initialisés aléatoirement) se caractérisent par des moyennes très semblables même si les regroupements diffèrent de taille. Ensuite, on observe que dans la comparaison des résultats des deux analyses par nuées dynamiques, les douze valeurs des centres de classes sont exactement les mêmes sept fois et semblables au centième près cinq fois. On considère donc ces ensembles de résultats comme pratiquement identiques.

2.3.5. Validation des résultats

La validité interne de la solution obtenue a été testée, en scindant au hasard l'échantillon initial ($n = 389$) en deux sous-échantillons ($n = 189$; $n = 200$). Les analyses ont été reprises avec chacun d'eux, comme le proposent Luke *et al*. (1991). Pour les deux sous-échantillons, les différences des moyennes avec l'échantillon initial sont au centième près.

La comparaison des moyennes de détresse psychologique entre les trois regroupements constitue un exemple de validation externe de la solution retenue. Les résultats d'une analyse de variance unifactorielle s'avèrent significatifs ($F = 54{,}23$; $dl = 2(386)$; $p < 0{,}001$) quant au niveau variable de détresse psychologique des trois regroupements.

Tableau 15.6.
Score moyen de chaque variable pour trois regroupements obtenus selon les différentes méthodes

Méthode	Variable	Regroupement 1		Regroupement 2		Regroupement 3	
		n	M	*n*	M	*n*	M
Hiérarchique[1]		198		123		68	
	TIF		2,16		3,61		3,83
	FIT		1,59		1,64		2,14
	ST		4,07		3,66		3,54
	SF		4,53		4,76		3,33
Par nuées dynamiques[2]		159		160		70	
	TIF		2,01		3,57		3,44
	FIT		1,49		1,75		2,09
	ST		4,06		3,72		3,66
	SF		4,69		4,69		3,03
Par nuées dynamiques[3]		159		160		70	
	TIF		2,01		3,57		3,45
	FIT		1,49		1,74		2,08
	ST		4,06		3,72		3,66
	SF		4,69		4,70		3,05

[1] Distance euclidienne au carré et méthode par agglomération utilisant l'algorithme de Ward.
[2] Centres de classes initiaux déterminés aléatoirement.
[3] Centres de classes initiaux obtenus de l'analyse hiérarchique.

2.3.6. Interprétation de la solution par nuées dynamiques

À partir des valeurs des moyennes aux quatre variables, les étiquettes données à chaque regroupement procèdent du construit d'équilibre (*balance*; Greenhaus et Allen, 2010) entre la vie professionnelle et la vie familiale. Le premier regroupement comprend 159 participants (40,9%) dont les scores aux deux variables d'interférence (TIF et FIT) sont inférieurs à la moyenne de l'échantillon de l'étude et les scores de satisfaction (ST et SF) sont supérieurs à la moyenne de l'échantillon (voir le tableau 15.5). Qualifié d'*équilibre global*, ce regroupement représente le rapport le plus harmonieux entre la vie professionnelle et la vie familiale. À la fois en déséquilibre et en équilibre, le deuxième regroupement se compose de 160 participants (41,1%) pour qui les scores TIF et ST sont respectivement supérieurs à la

moyenne et à la hauteur de celle-ci; en contrepartie, leurs scores FIT et SF sont respectivement inférieurs et supérieurs à la moyenne (figure 15.2). Un *équilibre partiel* caractérise ce regroupement, qui est partagé entre une vie professionnelle à satisfaction réduite qui interfère de façon très marquée avec les besoins de la famille et une vie familiale satisfaisante qui n'empiète pas sur le travail (figure 15.2). Une minorité de participants ($n = 70$; 18,0%) constitue le troisième regroupement, qui est atteint d'un *déséquilibre global*, car ses deux scores d'interférence (TIF et FIT) excèdent la moyenne alors que ses deux scores de satisfaction (ST et SF) sont inférieurs à celle-ci. Dans cette configuration de rapports discordants entre le travail et la famille et d'insatisfaction au sein de chacune de ces deux sphères de vie, le très bas niveau de satisfaction familiale est à retenir comme un facteur de risque supposément déterminant.

Comme il s'agit d'une première interprétation d'un tel ensemble de variables, des analyses prévues avec un autre échantillon comparable provenant d'une autre banque de données permettraient de vérifier la reproductibilité des présents résultats en termes de signification conceptuelle, du nombre de regroupements et du pourcentage de personnes qui les composent. Le cas échéant, des normes pourraient servir à soutenir les étiquettes données aux regroupements.

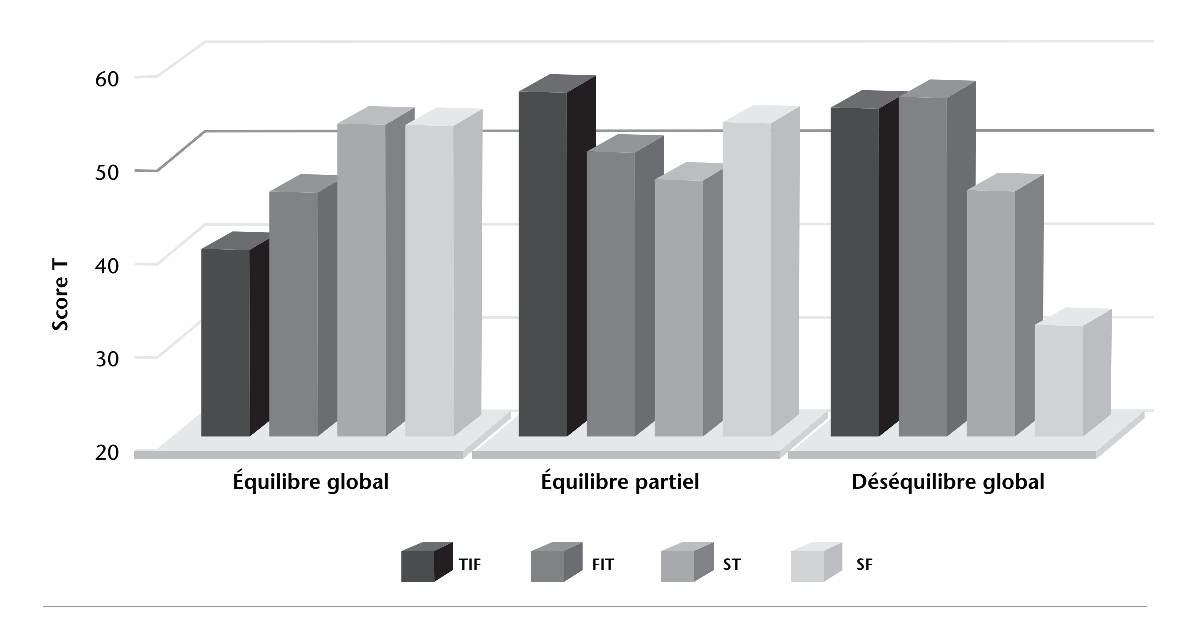
Afin de ramener à une même base de comparaison les résultats aux variables constituant les regroupements, celles-ci sont normalisées, puis graphiquement illustrées (tableau 15.7 et figure 15.2). Dans l'analyse de *Classification*, l'option *Tableau ANOVA* offerte par SPSS établit évidemment que, pour chaque variable, les tests F globaux d'analyses de variance sont tous significatifs, sans toutefois afficher de comparaisons de moyennes *a posteriori*. Pour observer les différences spécifiques de moyennes (< = >) entre les trois groupes à chacune des quatre variables, de telles comparaisons sont nécessaires. Elles montrent que les quatre variables ne distinguent pas de la même manière les trois regroupements de la typologie. Les variables TIF et SF avec les plus grands *F* sont les variables les plus discriminantes des regroupements entre eux.

Tableau 15.7.
Comparaison des scores normalisés (T*) aux variables d'interférence et de satisfaction des trois regroupements de la typologie

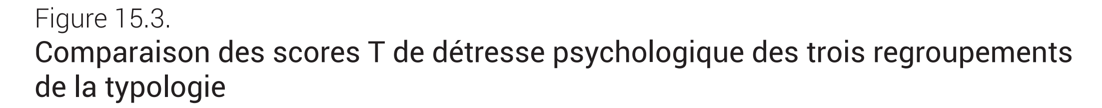
Variables	Équilibre global (*n* = 159)		Équilibre partiel (*n* = 160)		Déséquilibre global (*n* = 70)	*F***
TIF	40,33	<	57,13	=	55,67	360,35
FIT	46,30	<	50,75	<	56,70	31,25
ST	53,99	>	47,58	=	46,47	24,31
SF	53,91	=	53,88	>	32,26	433,93

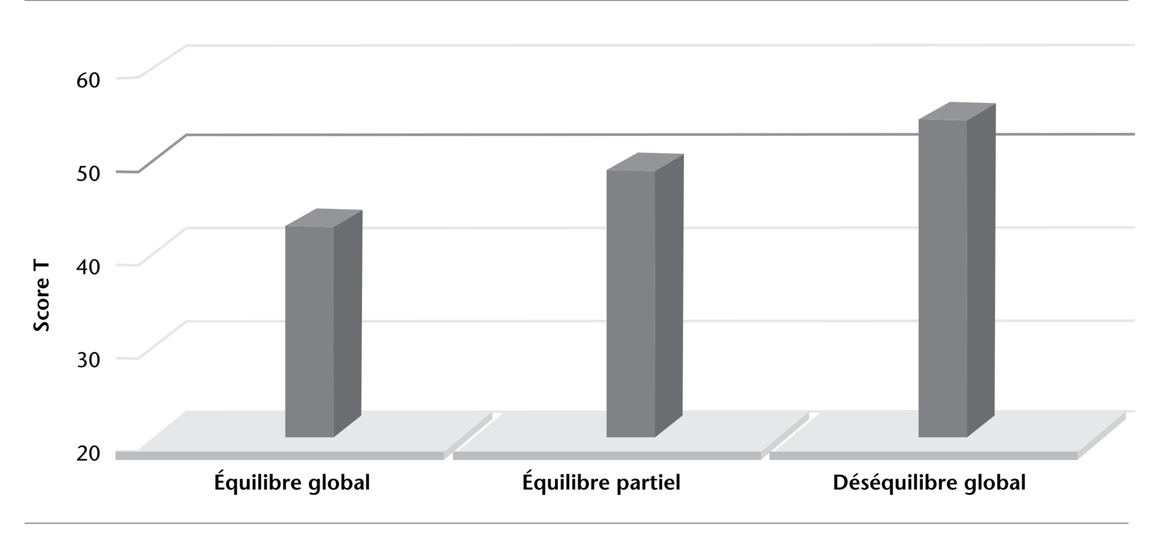
* T = (*z* * 10) + 50 ; score normalisé dont la moyenne = 50 et l'écart-type = 10.
** Pour toutes les valeurs de F, $p < 0{,}001$; *ddl* = 2(386).

Figure 15.2.
Score normalisé aux variables constituant les trois regroupements de la typologie



Les comparaisons de moyennes de détresse psychologique effectuées *a posteriori* (*Tukey B*) révèlent des différences de scores T entre chacun des trois regroupements (figure 15.3).


Figure 15.3.
Comparaison des scores T de détresse psychologique des trois regroupements de la typologie



Pour le *déséquilibre global*, dans le cas de 18 % des participants, le niveau élevé de détresse psychologique (*M* = 57,46 ; é.t. = 11,07) montre bien les répercussions conjointes des deux types d'interférence (TIF, FIT) sur les aspects négatifs de leur santé psychologique. Par ailleurs, selon la psychologie positive (Keyes et Lopez, 2002), le faible niveau de détresse psychologique de l'*équilibre global* (*M* = 45,11 ; é.t. = 7,60), qui caractérise 41 % des participants, ne suffit pas à lui seul à établir un bilan complet de leur santé psychologique. Pour y arriver, en plus d'une faible détresse psychologique, il serait nécessaire d'établir des mesures complémentaires de bien-être et de ressources personnelles.

Les résultats de notre étude permettent de répondre à la question initiale, à savoir que l'équilibre entre le travail et la famille est vécu différemment par divers groupes de personnes dont la santé psychologique se situe aussi à des niveaux variables. L'analyse de classification par regroupement a permis de constituer trois regroupements de personnes significativement distincts l'un de l'autre selon la qualité de leur équilibre travail-famille. Les tests F effectués *a posteriori* ont permis de comparer les trois groupes selon leurs différents niveaux de détresse psychologique.

CONCLUSION

Les approches d'analyse de données peuvent être centrées sur les variables ou sur les personnes. Dans le premier cas, à partir d'échantillons homogènes, l'attention se porte sur les relations entre les variables. Dans le second, on cherche plutôt à réunir en une catégorie des individus similaires entre eux (homogénéité intra-individuelle) et différents de ceux des autres catégories (hétérogénéité intergroupes). L'analyse de classification par regroupement appartient à cette approche centrée sur les personnes (Vondracek et Porfeli, 2004).

Dans le domaine de la santé psychologique et dans d'autres domaines connexes, cette méthode a comme avantage de définir, selon des variables précises, des ensembles distincts d'individus en vue de leur offrir des interventions, des services ou des soins différenciés et correspondant à leurs besoins particuliers. Différemment d'autres méthodes dont les conclusions se présentent sous forme de résultats de tests statistiques avec un niveau de signification bien précis, l'analyse de classification par regroupement est une méthode de type exploratoire. En ce sens, elle ne nécessite pas d'hypothèses *a priori*. Toutefois, elle implique un ensemble de décisions à prendre sur le plan théorique (variables pertinemment justifiées) et dans les étapes de sa réalisation (préparation ; exécution ; description, validation et interprétation des résultats).

Parmi les avantages de cette méthode, il y a le fait que les données sont regroupées de façon synthétique dans un nombre restreint de dimensions selon leur pertinence. Ainsi, les résultats sont présentés dans une solution parcimonieuse et facile à lire, même sous forme graphique (dendrogramme). Il est donc possible d'identifier des typologies qui ont des traits distinctifs sur la base des caractéristiques des données (cas) qui les constituent et des classes homogènes, dans lesquelles les unités constituantes sont interchangeables, et également de découvrir les liens entre les cas à l'intérieur des groupes homogènes (Green, Frank et Robinson, 1967).

La principale limite de l'analyse de classification par regroupement en est une de généralisation des résultats. En effet, il ne faut pas oublier que les individus d'un même regroupement ne sont semblables qu'en regard des informations (scores aux variables choisies) sur lesquelles l'analyse est basée. C'est pourquoi il est recommandé de scinder l'échantillon et d'appliquer la méthode deux fois. La méthode mixte (hiérarchique et par nuées dynamiques) se prête à tester la validité interne et externe. Dans notre exemple d'application, nous avons utilisé une variable externe (c.-à-d. détresse psychologique), mais la méthode permet d'en prendre d'autres. Il

s'avère aussi nécessaire de vérifier la reproductibilité des résultats et de leur interprétation en utilisant des données provenant d'une autre banque de données que celle utilisée dans une première analyse.

Dans une perspective exploratoire, les méthodes utilisées dans ce chapitre conviennent très bien au type de données habituellement traitées en psychologie. Toutefois, la prudence dans le choix d'une méthode serait de mise dans la poursuite d'objectifs différents ou propres à d'autres champs de recherche.

RÉFÉRENCES

ALDENDERFER, M.S. et R.K. BLASHFIELD (1984). *Cluster Analysis*, Beverly Hills, Sage Publications.

ANDES, N. (1998). « Robust strategies for application of cluster analysis », *Bulletin de méthodologie sociologique*, vol. 58, p. 31-47.

BLASHFIELD, R.K. et M.S. ALDENDERFER (1988). « The methods and problems of cluster analysis », dans J.R. Nesselroade et R.B. Cattell (dir.), *Handbook of Multivariate Experimental Psychology*, 2e éd., New York, Plenum Press, p. 447-473.

CARLSON, D.S. et J.G. GRZYWACZ (2008). « Reflections and future directions on measurement in work-family research », dans K. Korabik, D.S. Lero et D.L. Whitehead (dir.), *Handbook of Work-Family Integration : Research, Theory, and Best Practices*, New York, Academic Press, p. 57-73.

DEMEROUTI, E. et S. GEURTS (2004). « Towards a typology of work-home interaction », *Community, Work & Family*, vol. 7, p. 285-309.

DORIO, J.M., R.H. BRYANT et T.D. ALLEN (2008). « Work-related outcomes of work-family interface : Why organizations should care », dans K. Korabik, D.S. Lero et D.L. Whitehead (dir.), *Handbook of Work-Family Integration : Research, Theory, and Best Practices*, New York, Academic Press, p. 157-176.

EVERITT, B., S. LANDAU et M. LEESE (2001). *Cluster Analysis*, 4e éd., Londres, Arnold.

FOURNIER, S. (1995). *Relations entre la congruence individu-environnement, l'importance du groupe de travail et la satisfaction au travail*, Mémoire de maîtrise, Montréal, Université de Montréal.

FRONE, M.R., M. RUSSELL et M.L. COOPER (1992). « Antecedents and outcomes of of work-family conflict : Testing a model of the work-family interface », *Journal of Applied Psychology*, vol. 77, p. 65-78.

GREEN, P.E., R.E. FRANK et P.J. ROBINSON (1967). « Cluster analysis in text market selection », *Management Science*, avril, p. 387-400.

GREENHAUS, J.H. et T.D. ALLEN (2010). « Work-family balance : A review and extension of the literature », dans J.C. Quick et L.E. Tetrick (dir.), *Handbook of Occupational Health Psychology*, 2e éd., Washington, D.C., American Psychological Association, p. 165-183.

Halkidi, M., Y. Batistakis et M. Vazirgiannis (2001). « On clustering validation techniques », *Journal of Intelligent Information Systems*, vol. 17, nos 2-3, p. 107-145.

Hennig, C. (2008). « Dissolution point and isolation robustness : Robustness criteria for general cluster analysis methods », *Journal of Multivariate Analysis*, vol. 99, no 6, p. 1154-1176.

Henry, D.B., P.H. Tolan et D. Gorman-Smith (2005). « Clustering methods in family psychology research », *Journal of Family Psychology*, vol. 19, p. 121-132.

IBM Corporation (2012). *IBM SPSS Statistics Base 21*, <ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/21.0/fr/client/Manuals/IBM_SPSS_Statistics_Base.pdf>, consulté le 3 décembre 2012.

Jain, A.K., M.N. Murty et P.J. Flynn (2000). « Data clustering : A review », *ACM Computing Surveys*, vol. 31, no 3, p. 264-323.

Khalid, D. et A. Bouklouze (2006). « Outils statistiques pour l'analyse », *Les technologies de laboratoire*, vol. 1, p. 15-19.

Keyes, C.L.M. et S.J. Lopez (2002). « Toward a science of mental health : Positive directions in diagnosis and interventions », dans C.R. Snyder et S.J. Lopez (dir.), *Handbook of Positive Psychology*, New York, Oxford University Press, p. 45-59.

Kinnunen, U., S. Geurts et S. Mauno (2004). « Work-to-family conflict and its relationship with satisfaction and well-being : A one-year longitudinal study on gender differences », *Work & Stress*, vol. 18, p. 1-22.

Kopelman, R.E., J.H. Greenhaus et T.F. Connolly (1983). « A model of work, family, and inter-role conflict : A construct-validation study », *Organizational Behavior and Human Performance*, vol. 32, p. 198-215.

Légaré, G., M. Préville, R. Massé, C. Poulin, D. St-Laurent et R. Boyer (2000). « Santé mentale », dans Santé Québec (dir.), *Enquête sociale et de santé 1998*, Québec, Institut de la statistique du Québec, p. 333-352.

Lorr, M. (1983). *Cluster Analysis for Social Scientists*, San Francisco, Jossey-Bass.

Luke, D.A., J. Rappaport et E. Seidman (1991). « Setting phenotypes in a mutual help organization : Expanding behavior setting theory », *American Journal of Community Psychology*, vol. 19, p. 147-167.

Milligan, G.W. (1981). « A Review of Monte Carlo Tests of Cluster Analysis », *Multivariate Behavioral Research*, vol. 16, 379-407.

Milligan, G.W. et S.C. Hirtle (2003). « Clustering and classification methods », dans J. Schinka et W. Velicer (dir.). *Comprehensive Handbook of Psychology,* vol. 2, New York, Wiley & Sons, p. 165-186.

Moran, C.M., J.M. Diefendorff, T.-Y. Kim et Z.-Q. Liu (2012). « A profile approach to self-determination theory motivations at work », *Journal of Vocational Behavior*, vol. 81, p. 354-363.

Mullen, J., E. Kelley et E.K. Kelloway (2008). « Health and well-being outcomes of the work-family interface », dans K. Korabik, D.S. Lero et D.L. Whitehead (dir.), *Handbook of Work-Family Integration : Research, Theory, and Best Practices*, New York, Academic Press, p. 191-214.

Parasuraman, S., Y.S. Purohit, V.M. Godshalk et N.J. Beutell (1996). « Work and family variables, entrepreneurial career success, and psychological well-being », *Journal of Vocational Behavior*, vol. 48, p. 275-300.

Perron, J. (1999). *Analyse de la culture d'une entreprise pharmaceutique*, Rapport de recherche inédit, Montréal, Université de Montréal.

Perron, J., P. Lamothe et J.-P. Daoust (2005). « Bien-être subjectif et santé psychologique : profils du personnel d'établissements d'éducation », dans A. Battistelli, M. Depolo et F. Fraccaroli (dir.), *La qualité de vie au travail dans les années 2000*, Actes du 13e Congrès de psychologie du travail et des organisations, Bologne, CLUEB, p. 1268-1274.

Rantanen, J., L. Pulkkinen et U. Kinnunen (2005). « The big five personality dimensions, work-family conflict, and psychological distress : A longitudinal view », *Journal of Individual Differences*, vol. 26, p. 155-166.

Rapkin, B.D. et D.A. Luke (1993). « Cluster analysis in community research : Epistemology and practice », *American Journal of Community Research*, vol. 21, p. 247-277.

Rujasiri, P., et B. Chomtee (2009). « Comparison of clustering techniques for cluster analysis », *Kasetsart J. (Nat. Sci.)*, vol. 43, p. 378-388.

Saitta, S., B. Raphael et I.F.C. Smith (2008). « A comprehensive validity index for clustering », *Intelligent Data Analysis*, vol. 12, nº 6, p. 529-548.

Schwarz, G.E. (1978). « Estimating the dimension of a model », *Annals of Statistics*, vol. 6, nº 2, p. 461-464, <doi:10.1214/aos/1176344136>.

Scoltock, J. (1982). « A survey of the literature of cluster analysis », *Computer Journal*, vol. 25, nº 1, p. 130-134.

Sokal, R.R. et P.H.A. Sneath (1963). *Principles of Numerical Taxonomy*, San Francisco, W.H. Freeman.

Steinley, D. (2006). « K-means clustering : A half-century synthesis », *British Journal of Mathematical and Statistical Psychology*, vol. 59, nº 1, p. 1-34, <http://dx.doi.org/10.1348/000711005X48266>.

Tryon, R.C. (1955). *Identification of Social Areas by Cluster Analysis*, Berkeley, University of California Press.

Tryon, R.C. et D.E. Bailey (1970). *Cluster Analysis*, New York, McGraw-Hill.

Vondracek, F. et E.J. Porfeli (2004). « Perspectives historiques et contemporaines sur la transition de l'école au travail : apports théoriques et méthodologiques du modèle développemental-contextuel », *L'orientation scolaire et professionnelle*, vol. 33, nº 3, p. 1-19.

Ward, J. (1963). « Hierarchical grouping to optimize an objective function », *Journal of the American Statistical Association*, vol. 58, p. 236-244.

Weiss, D.J., R.V. Dawis, G.W. England et L.H. Lofquist (1967). *Manual for the Minnesota Satisfaction Questionnaire*, St. Paul, Work Adjustment Project, Industrial Relations Centre, University of Minnesota.

CHAPITRE 16

ANALYSES DE VARIANCE UNIVARIÉE ET MULTIVARIÉE

Illustration de l'analyse des problèmes de comportement chez les jeunes ayant une déficience intellectuelle selon le sexe et le niveau de déficience

Lise Lachance
Gilles Raîche

FORCES

- Les analyses de variance univariée (ANOVA) et multivariée (MANOVA) sont particulièrement utiles dans le contexte de devis de recherche expérimentaux ou quasi expérimentaux au sein desquels le chercheur souhaite comparer l'effet de traitements sur une ou des variables dépendantes quantitatives.
- Elles permettent de tester simultanément l'effet de plusieurs variables indépendantes qualitatives ainsi que de leur interaction sur la ou les variables dépendantes quantitatives.
- Elles peuvent servir à comparer des moyennes obtenues dans différentes conditions (expérimentales ou non) auprès des mêmes individus ou non (variables à mesures répétées ou à groupes indépendants) en contrôlant ou non l'effet de certaines variables confondantes (covariables).

LIMITES

- L'ANOVA et la MANOVA sont des méthodes d'analyse sensibles aux valeurs extrêmes des variables dépendantes, lesquelles peuvent avoir un effet sur la probabilité d'erreur de type I.
- La puissance statistique de l'ANOVA et de la MANOVA est plus faible lorsque la taille des groupes est très inégale.
- Ces deux types d'analyse sont difficilement applicables lorsqu'il y a des données manquantes dans les devis à mesures répétées. C'est notamment pour cette raison que la modélisation à niveaux multiples (voir le chapitre 20 de cet ouvrage) tend de plus en plus à remplacer la MANOVA.

L'analyse de variance tire principalement son origine des travaux de Fisher (1925, 1935, 1956) dans le contexte du développement des plans expérimentaux et de l'expérimentation en agriculture ainsi que du développement de la théorie fréquentiste (ou classique) des tests d'hypothèses (Lehmann, 2011), théorie qui s'oppose à l'approche dite subjectiviste propre aux modèles bayésiens (Robert, 2007). L'analyse de variance est souvent appelée ANOVA en raison du terme anglais (***AN**alysis **O**f **VA**riance*). Il s'agit d'une méthode d'analyse qui vise à déterminer si une ou des variables indépendantes qualitatives[1] (appelées aussi facteurs[2] ou traitements) ont des effets significatifs sur une variable dépendante quantitative[3]. Quand plusieurs variables indépendantes sont considérées simultanément, l'ANOVA permet également d'évaluer si leur effet d'interaction explique significativement le résultat obtenu à la variable dépendante. Comme l'ANOVA porte sur une seule variable dépendante, elle fait partie des méthodes d'analyse univariées. Contrairement au test t, elle permet de comparer plus de deux moyennes à la fois. Lorsque deux moyennes sont comparées, l'ANOVA donne les mêmes résultats qu'un test t pour des échantillons indépendants (deux ensembles d'observations différentes) ou qu'un test t pour échantillons pairés (un même ensemble d'observations qui répond à deux reprises ou deux ensembles d'observations reliées). Cependant, les calculs pour effectuer un test t sont plus simples que ceux liés à une ANOVA. D'ailleurs, lorsqu'il y a seulement deux groupes à comparer, la statistique F obtenue à l'ANOVA est égale au score t mis au carré. Bien sûr, à ce moment, le signe de la différence est perdu.

L'objectif poursuivi dans ce chapitre est de donner un aperçu de deux techniques de comparaison de moyennes : l'ANOVA et la MANOVA. Dans un premier temps, il y sera question des différents plans d'analyse de même que de la logique sous-jacente à l'ANOVA et des particularités de la MANOVA. Puis, leurs principales conditions d'application seront abordées ainsi que des transformations de données susceptibles d'aider à les remplir lorsque ce n'est pas le cas. Les étapes importantes de la démarche d'analyse de l'ANOVA ainsi que celles de la MANOVA seront ensuite traitées.

---

1. Plus précisément, il est question ici d'une variable discrète qui peut être représentée sur une échelle nominale ou ordinale.
2. Il ne faut pas confondre le concept de facteur dans une ANOVA avec celui utilisé dans le contexte d'une analyse factorielle exploratoire ou confirmatoire (voir le chapitre 22 de cet ouvrage). Dans ce dernier cas, il s'agit plutôt d'une variable latente qui peut être mesurée à partir de variables manifestes ou observables.
3. Il arrive fréquemment qu'on utilise des variables catégorielles ordonnées (discrètes) comme des variables quantitatives (continues). Il n'y a pas de règle formelle quant au nombre de catégories ordonnées nécessaires pour fournir une approximation satisfaisante d'une variable quantitative. Le nombre de sept a souvent été proposé, mais lorsque le coefficient d'asymétrie est élevé, ce nombre n'est pas approprié. Les méthodes d'analyse modernes qui prennent en compte les variables catégorielles ordonnées dépassent toutefois le propos de ce chapitre. Elles relèvent généralement des modélisations issues de la théorie de la réponse à l'item (Thissen et Wainer, 2001) ou de l'étalonnage multidimensionnel (Nishisato, 2004).

Le chapitre se termine avec un exemple d'application de la MANOVA, suivie d'ANOVA. Les données utilisées à cette fin sont réelles[4] et proviennent d'un projet de recherche mené par Lachance, Richer, Côté et Tremblay (2010).

1. PLANS D'ANALYSE DE VARIANCE

Selon le nombre de variables indépendantes à l'étude, le plan d'ANOVA peut être qualifié de simple (une seule variable) ou de factoriel (plus d'une variable). De plus, selon la nature des observations recueillies, les facteurs peuvent être à groupes indépendants (un participant est soumis à une seule condition), à mesures répétées (chaque participant est soumis à toutes les conditions) ou mixtes, c'est-à-dire les deux à la fois. Dans le cas des groupes indépendants, on s'intéresse alors à la variabilité intergroupes, tandis qu'avec les mesures répétées, c'est à la variabilité intrasujet à laquelle on porte intérêt.

Par convention, les plans factoriels d'ANOVA peuvent être représentés en considérant le nombre de niveaux (ou de catégories) de chacun des facteurs (ou variables indépendantes). De plus, l'utilisation de parenthèses sert à identifier les niveaux des facteurs à mesures répétées. Par exemple, si le plan d'ANOVA est dit 2×4, on peut en déduire qu'il y a 8 groupes indépendants, formés à partir de 2 facteurs. Le premier contient 2 niveaux (p. ex., sexe : féminin et masculin) alors que le second en inclut 4 (p. ex., niveau de scolarité : primaire, secondaire, collégial, universitaire). De plus, une ANOVA $2 \times 2 \times (3)$ désigne un plan mixte à trois facteurs, c'est-à-dire à 4 groupes indépendants et à mesures répétées sur le troisième.

Un facteur est dit aléatoire si les niveaux (ou les catégories) des variables indépendantes à l'étude ont été choisis au hasard parmi l'ensemble des niveaux possibles. Si les niveaux ont été définis par le chercheur en fonction de ses intérêts, le facteur est alors qualifié de fixe.

2. LOGIQUE DE L'ANALYSE

De prime abord, il peut sembler étrange qu'une méthode d'analyse destinée à comparer des moyennes soit appelée analyse de variance. Or, cette appellation vient du fait que pour tester globalement la différence entre plusieurs

4. Contrairement aux données présentées dans le rapport de recherche, les jeunes ne répondant pas au critère diagnostique relatif au fonctionnement adaptatif de la déficience intellectuelle n'ont pas été exclus de l'échantillon afin d'en maximiser la taille aux fins de l'exemple.

moyennes, il faut analyser les variances. En effet, si les moyennes de chacun des groupes à l'étude ne diffèrent pas statistiquement entre elles, leur variance devrait s'approcher de la valeur 0.

De façon plus spécifique, l'objectif poursuivi par l'ANOVA consiste à expliquer à quoi est attribuable la variance observée sur une variable dépendante ou à en déterminer les différentes causes possibles. Deux sources principales peuvent servir à cette fin : la variance expliquée par un facteur ou l'appartenance à un groupe (variabilité entre les groupes ou intergroupes) et celle expliquée par d'autres facteurs non contrôlés ou considérés dans l'étude qui sont, par conséquent, vus comme de l'erreur aléatoire (variabilité entre les individus d'un groupe ou intragroupe). Cette dernière inclut à la fois les différences individuelles (p. ex., habiletés personnelles, motivation) et l'erreur expérimentale (p. ex., consignes données aux participants de façon non uniforme, différences dans la calibration de l'équipement). C'est pourquoi la variabilité intragroupe est aussi appelée variabilité résiduelle.

Même si les groupes à l'étude provenaient de la même population, il serait peu probable que ceux-ci présentent exactement la même moyenne en raison des fluctuations liées à l'échantillonnage. L'ANOVA vise donc à tester si les différences observées sont attribuables à la variabilité naturelle de la moyenne entre différents échantillons d'une même population ou bien s'il y a une raison de croire qu'il existe un ou des groupes qui se distinguent réellement de la moyenne populationnelle.

Dans les grandes lignes, voici les étapes de l'ANOVA. D'abord, la moyenne générale est calculée en faisant la moyenne de toutes les observations (ou lorsque les groupes sont égaux, la moyenne des moyennes des groupes). Les moyennes de chacun des groupes sont ensuite évaluées. La variance intergroupes est déterminée en établissant les écarts entre les moyennes des groupes et la moyenne générale. Par la suite, la variance au sein de chaque groupe est calculée en considérant l'écart entre les scores individuels et la moyenne de son groupe respectif. Comme plusieurs estimés de la variance intragroupe sont disponibles, il suffit d'en faire la moyenne pour en obtenir un seul, lequel est un estimé de la variance de la population. En somme, ces calculs permettent de découper la variance totale en deux parties : la variance intergroupes et la variance intragroupe.

L'hypothèse nulle de l'ANOVA (à l'aide du test F) suppose que toutes les moyennes sont égales (ou semblables), car elles proviennent de la même population, alors que l'hypothèse alternative stipule qu'au moins une des moyennes est différente des autres. Dans ce contexte, le test d'hypothèses associé ne tient pas compte de la direction des différences observées entre les moyennes et peut ainsi être considéré comme un test non directionnel.

Si l'hypothèse nulle est vraie, cela implique que les différences observées entre les moyennes des groupes représentent de l'erreur d'échantillonnage liée aux différences individuelles. Dans ce cas, la variance intergroupes peut être considérée comme un autre estimé de la variance de la population.

Le test F permet de vérifier l'hypothèse nulle, qui suppose l'égalité entre les moyennes au sein de la population étudiée, à partir du rapport entre la variabilité intergroupes et la variabilité intragroupe. Comme la variabilité intergroupes est le numérateur de ce rapport, plus les moyennes sont éloignées les unes des autres, plus la valeur F est élevée. Il devient donc peu probable que celle-ci soit attribuable uniquement à l'erreur d'échantillonnage. Pour déterminer s'il existe une différence significative entre les groupes et vérifier que les écarts ne sont pas dus à une fluctuation liée à l'échantillonnage, il faut comparer la valeur F obtenue à une distribution d'échantillonnage connue : la distribution F. Le degré de signification du test dépendra de trois facteurs : la valeur F et les deux degrés de liberté (intergroupe au numérateur et intragroupe au dénominateur).

Si l'hypothèse nulle n'est pas rejetée, cela signifie que les différences observées sur la variable dépendante ne sont pas suffisantes pour juger qu'elles relèvent d'autres facteurs que de l'erreur d'échantillonnage. Conséquemment, les moyennes des groupes ne peuvent être considérées comme différentes. La logique de l'approche classique des tests d'hypothèses suggère alors non pas d'accepter l'hypothèse nulle, mais plutôt de suspendre notre jugement. Si l'hypothèse nulle est rejetée et que le test est dit significatif, cela implique que la variation totale à la variable dépendante peut être expliquée par la variance entre les groupes et que cette dernière est supérieure à celle expliquée par l'erreur d'échantillonnage ou à l'intérieur des groupes.

En somme, l'ANOVA permet de départager la variance explicable par un facteur de la variance inexpliquée (variance intragroupe ou résiduelle). Conséquemment, l'ajout de facteurs dans une ANOVA vise à réduire la portion de variance inexpliquée par la prise en compte de ceux-ci et permet, du même coup, d'analyser l'impact combiné de plusieurs variables indépendantes. En plus d'ajouter des facteurs supplémentaires dans l'analyse, le chercheur peut y introduire des covariables pour réduire la part de la variance inexpliquée ou résiduelle. Il s'agit de variables quantitatives susceptibles de « covarier » avec la variable dépendante et d'affecter indirectement la relation qui existe entre la ou les variables indépendantes et la variable dépendante. Autrement dit, puisque les unités d'observation ne sont pas attribuées au hasard aux différentes valeurs possibles des variables confondantes, il s'agit d'un moyen statistique, plutôt qu'expérimental, de contrôler l'effet de variables confondantes. Il faut toutefois noter qu'en général il vaut

mieux exercer un contrôle expérimental qu'un contrôle statistique à l'aide de variables confondantes lorsque c'est possible, car le contrôle statistique repose sur le respect de certains postulats. Quand une covariable est intégrée dans l'analyse, il est alors question d'ANCOVA (pour *Analysis of Covariance*).

En conclusion, il importe de souligner que l'ANOVA peut être aussi considérée sous l'angle des modèles linéaires généraux (Cohen et Cohen, 1983; Searle, 1971) et ainsi comme un simple modèle de régression multiple. Il ne faut cependant pas confondre les modèles linéaires généraux avec les modèles linéaires généralisés (McCullagh et Nelder, 1989), qui sont en fait des modèles non linéaires qui peuvent être linéarisés par une transformation exponentielle de la relation entre la variable dépendante et les variables indépendantes. Dans ce modèle, une variable dépendante y est expliquée par p variables indépendantes quantitatives continues $\{x_1$ à $x_p\}$ auxquelles sont associés p coefficients de régression multiple $\{b_1$ à $b_p\}$.

$$y = b_0 + b_1x_1 + b_2x_2 + \ldots + b_px_p + e$$

Dans l'ANOVA, les variables indépendantes quantitatives continues sont remplacées par k – 1 variables catégorielles où k correspond au nombre de catégories de la variable catégorielle. Cette contrainte est nécessaire, car la valeur obtenue à la k^e variable catégorielle est totalement déterminée par celle des k – 1 autres catégories. Selon l'objectif de comparaison des moyennes, il y a plusieurs façons de coder chacune des catégories de la variable indépendante, mais le résultat aux tests de signification statistiques demeure le même.

3. PARTICULARITÉS DE L'ANALYSE DE VARIANCE MULTIVARIÉE

Le concept d'analyse de variance multivariée (MANOVA pour ***M**ultivariate* ***AN**alysis* ***O**f* ***VA**riance*) est une extension de l'ANOVA qui est apparue parallèlement avec les travaux d'Hotelling (1931, 1936) sur une extension multivariée du test *t,* soit le T^2, et sur l'analyse par corrélations canoniques ainsi que ceux de Wishart (1928) sur les matrices de covariances et de Wilks (1932) sur une statistique multivariée analogue à la variance, soit le lambda de Wilks. Toutefois, en raison de la complexité des calculs se rattachant à la MANOVA et du manque de programmes informatiques disponibles, son application dans le domaine des sciences humaines et sociales a tardé (Bray et Maxwell, 1982; Hair, Black, Babin et Anderson, 2010).

En fait, la MANOVA permet d'évaluer l'effet d'une ou de plusieurs variables indépendantes qualitatives[5] sur un ensemble de variables dépendantes quantitatives simultanément. Autrement dit, elle vise à expliquer à quoi est attribuable la variance observée sur un ensemble de variables dépendantes. Par l'analyse de multiples variables dépendantes à la fois, la MANOVA offre l'avantage de mieux contrôler la probabilité d'erreur de type I (c'est-à-dire de rejeter l'hypothèse nulle alors qu'elle est vraie ou de conclure qu'il y a une différence significative, alors que ce n'est pas le cas dans la réalité). En effet, ce risque d'erreur augmente notamment en fonction du nombre de tests réalisés auprès des mêmes participants ainsi que du degré de corrélation entre les variables dépendantes (Cole, Maxwell, Arvey et Salas, 1994). Avec la MANOVA, on ne fait alors qu'un seul test d'hypothèses sur l'ensemble des variables dépendantes plutôt qu'un test d'hypothèses par variables dépendantes. C'est ce qui permet de mieux contrôler l'erreur de type I. De manière corollaire, comparativement à l'ANOVA, la MANOVA permet d'accroître la puissance statistique. À ce propos, celle-ci peut parvenir à détecter globalement une réelle différence significative entre les groupes qui serait trop petite pour être décelée par une ANOVA réalisée sur l'une ou l'autre des variables séparément.

Il faut toutefois noter que, s'il y a de fortes corrélations entre les variables, une variable dépendante peut n'être qu'une combinaison linéaire des autres variables dépendantes. On dit alors qu'il y a colinéarité entre les variables. Dans ces circonstances, il peut devenir redondant sur le plan statistique d'inclure certaines variables. C'est d'ailleurs pour cette raison qu'un résultat global à un test n'est pas analysé en même temps que les résultats obtenus aux sous-échelles de celui-ci.

Rappelons que l'hypothèse nulle de l'ANOVA stipule que les moyennes de tous les groupes sont égales et qu'elles proviennent toutes de la même population. Comme plusieurs variables dépendantes sont considérées dans la MANOVA, il est plutôt question de vecteurs de moyennes. Ceux-ci sont formés par la combinaison linéaire des variables dépendantes afin d'obtenir une ou plusieurs variables « synthétiques », dites discriminantes, qui permettent d'obtenir une réduction dimensionnelle et qui serviront aux fins d'analyse. Ainsi, l'hypothèse nulle de la MANOVA renvoie à l'égalité des vecteurs de moyennes.

Bien que cela puisse paraître abstrait, il n'est pas rare que les chercheurs utilisent les combinaisons linéaires. C'est le cas lorsqu'ils combinent les scores obtenus à plusieurs sous-échelles ou sous-tests pour former un score

5. Plus précisément, il est question ici d'une variable discrète qui peut être représentée sur une échelle nominale ou ordinale.

global. À titre d'exemple, Blais, Sabourin, Boucher et Vallerand (1990) ont utilisé une formule pour estimer le niveau d'autodétermination des individus en combinant les résultats qu'ils ont obtenus aux diverses sous-échelles d'un questionnaire de motivation : *Autodétermination = [(3*motivation intrinsèque) + (2*régulation intégrée) + (1*régulation identifiée)] – [(1*régulation introjectée) + (2*régulation externe) + (3*amotivation)].*

La question qui demeure dans le contexte de la MANOVA concerne la façon dont la ou les combinaisons linéaires les plus appropriées sont choisies parmi l'ensemble des combinaisons possibles. En fait, celles retenues ont pour but de maximiser la différence entre les groupes. Lorsqu'il y a plus de deux groupes, plusieurs combinaisons linéaires peuvent être utiles pour discriminer entre ceux-ci. Ces combinaisons linéaires sont alors nommées des fonctions discriminantes.

Plusieurs méthodes de calcul peuvent être appliquées pour déterminer la meilleure combinaison linéaire de variables dépendantes. Les plus fréquentes sont le lambda de Wilks, la trace de Pillai, la trace de Hotelling et la plus grande valeur propre de Roy. La signification de ces quatre statistiques est vérifiée à l'aide d'une approximation du test F (Bock, 1975 ; Fox, 2008 ; SAS Institute, 1990). Comme l'indiquent les équations ci-dessous, ces statistiques sont toutes basées sur le calcul des valeurs propres L_j du rapport de la matrice des produits croisés de l'hypothèse sur la matrice des produits croisés des résidus ($SSP_H SSP_R^{-1}$).

Trace de Pillai $$V = \sum_{j=1}^{m} \frac{L_j}{1 - L_j}$$

Trace de Hotelling $$U = \sum_{j=1}^{m} L_j$$

Plus grande valeur propre de Roy $$\Theta = L_1$$

Lambda de Wilks $$A = \sum_{j=1}^{m} \frac{L_j}{1 + L_j}$$

En pratique, les différentes valeurs de *F* résultant de ces statistiques sont souvent égales[6] ou très près les unes des autres (Tabachnik et Fidell, 2013). Ainsi, dans la plupart des situations, les conclusions statistiques découlant de l'application de ces tests convergent. Il arrive cependant qu'elles diffèrent et, dans ce cas, la question est de déterminer laquelle est

6. C'est le cas lorsqu'il n'y a que deux groupes à comparer.

la plus pertinente et devrait être retenue étant donné la distribution des données recueillies. Sur cet aspect, il ne semble pas y avoir de consensus dans les écrits. Or, selon Tabachnik et Fidell (2013), le lambda de Wilks semble s'imposer, quoique d'autres préfèrent la trace de Pillai, car celle-ci offrirait une plus grande puissance et serait plus robuste lorsque le postulat d'homogénéité de la variance n'est pas respecté (Olson, 1976).

4. CONDITIONS D'APPLICATION LIÉES À L'ANOVA ET À LA MANOVA

Les méthodes d'analyse paramétriques, qui ont pour but de faire une inférence statistique à propos des paramètres de la population à partir d'échantillons, reposent sur un ensemble de conditions d'application appelées postulats. Ces conditions devraient être vérifiées et idéalement satisfaites pour nous permettre d'interpréter adéquatement les résultats obtenus à ces méthodes d'analyse. Il s'agit principalement de l'absence de valeurs extrêmes ou influentes, de la complétude des données, de la linéarité des relations, de l'indépendance des observations, de la normalité de la distribution, de l'homogénéité de la variance et de celle des matrices de covariances.

4.1. Absence de valeurs extrêmes ou influentes

Comme le calcul de la moyenne est affecté par des valeurs extrêmes ou influentes, il est peu surprenant que les résultats des méthodes ayant pour but la comparaison de moyennes, comme l'ANOVA et la MANOVA, le soient également. Il devient donc important d'analyser les observations atypiques qui peuvent en fausser les résultats. Cependant, il faut noter que celles-ci peuvent être extrêmes sans pour autant être influentes. Bien que l'on puisse être tenté d'éliminer ces observations, il ne faut pas oublier que leur retrait peut ne pas avoir d'impact sur l'estimation des paramètres et qu'elles peuvent être strictement issues de la population à l'étude. Il n'est donc pas toujours pertinent de les retirer de l'analyse.

Les valeurs extrêmes ou influentes univariées proviennent de participants qui présentent un résultat aberrant sur une variable alors que les valeurs extrêmes ou influentes multivariées concernent une combinaison inhabituelle de résultats sur plusieurs variables (p. ex., être très jeune et avoir un revenu très élevé). Les valeurs extrêmes ou influentes univariées sont généralement déterminées par le calcul d'une statistique dite de distance à l'aide de scores standardisés (scores z) ou par le diagramme en boîte et moustaches (*box plot*).

Après avoir transformé les résultats en scores *z*, les participants qui obtiennent des valeurs inférieures à –3,00 ou supérieures à 3,00 peuvent être considérés potentiellement comme des valeurs extrêmes ou influentes. En effet, selon la table de distribution normale, la probabilité d'obtenir de telles valeurs est inférieure à 0,001, soit à une chance sur 1 000.

Le diagramme en boîte et moustaches est une représentation graphique très pratique pour illustrer la distribution empirique des données par le biais de quelques mesures de position : la médiane (M), le premier (*Q*1) et le troisième quartile (*Q*3) (voir la figure 16.1 dans l'exemple à la fin du chapitre). Comme les extrémités des moustaches sont délimitées par 1,5 fois l'intervalle interquartile (*Q*3 – *Q*1), il permet de déceler la présence de valeurs extrêmes en plus de vérifier la normalité de la distribution. Ce graphique est souvent préféré au calcul des scores *z*, car il ne repose pas sur la moyenne et la symétrie de la distribution. Il dépend des mesures de position, qui sont peu influencées par les valeurs extrêmes.

Pour ce qui est de l'identification des valeurs extrêmes ou influentes multivariées, plusieurs méthodes peuvent être employées. Celles qu'on retrouve le plus fréquemment sont la distance, la force de levier (*leverage*) et l'influence (Fox, 2008 ; Howell, 2008).

La distance concerne strictement les variables dépendantes et est généralement calculée à l'aide de la distance de Mahalanobis (Anderson, 2003). Elle représente la distance entre une observation et la moyenne de toutes les autres variables dépendantes. Afin d'établir à partir de quelle valeur une observation peut être jugée atypique, Barnett et Lewis (1978) ont produit une table de valeurs critiques qui tient compte du nombre de variables et de la taille de l'échantillon. Pour leur part, Tabachnick et Fidell (2013) suggèrent de comparer la distance de Mahalanobis obtenue pour chaque observation à la valeur critique de la table de χ^2, associée à une probabilité inférieure à 0,001 et dont les degrés de liberté correspondent au nombre de variables dépendantes.

La force de levier (h_i ou *hat value*) sert à identifier des observations qui se démarquent de manière importante des variables indépendantes (Faraway, 2005 ; Howell, 2008). Son calcul nécessite toutefois de recourir au calcul matriciel. Puisque $\sum h_i = p$, la moyenne de la force de levier est égale à p/n où p représente le nombre de variables indépendantes et n le nombre d'observation. En général, il est suggéré de porter une attention particulière aux valeurs supérieures à $2p/n$ (Faraway, 2005) ou encore à $3(p + 1)/n$ (Stevens, 2009).

Enfin, l'influence permet de combiner simultanément la distance et la force de levier. Le D de Cook est la statistique d'influence qui est sûrement la plus fréquemment utilisée (Faraway, 2005). Elle représente la différence entre les résultats actuels et ceux qui auraient été obtenus si l'observation avait été éliminée de l'analyse. Selon Howel (2008), il n'existe pas de norme reconnue pour déterminer la valeur critique de D. D'après Cook et Weisberg (1982), les observations ayant des valeurs supérieures à 1 peuvent être jugées atypiques.

Dans le cas de petits échantillons, il importe de porter une attention particulière à l'impact des valeurs extrêmes ou influentes. Il est également recommandé de faire l'analyse avec ou sans ces valeurs pour voir si les résultats des deux analyses convergent (Kruskal, 1960). Si les résultats convergent, il n'y a donc pas lieu de s'inquiéter outre mesure concernant celles-ci. Plusieurs autres stratégies peuvent être adoptées pour réduire l'influence des valeurs extrêmes sur les résultats : l'élimination de l'observation, le remplacement de celles-ci par la moyenne des autres observations, etc. Cependant, il ne faut pas oublier que ces stratégies comportent le risque de fausser les résultats lorsque ces valeurs existent au sein de la population.

4.2. Complétude des données

Il arrive fréquemment que les données disponibles comportent des données manquantes. Dans ce cas, quelle que soit la méthode d'analyse utilisée, l'interprétation des résultats en est affectée. Traditionnellement, on a suggéré de retirer l'observation en entier (*listwise*), de traiter les observations deux à deux et de considérer l'ensemble des paires d'observations disponible (*pairwise*) ou de remplacer les valeurs manquantes par une mesure de tendance centrale (moyenne, médiane, etc.). Ce sont d'ailleurs ces approches qui sont les plus faciles à utiliser dans les logiciels courants. Toutefois, lorsque le nombre de données manquantes est important, à tout le moins supérieur à 5 %, celles-ci produisent des estimateurs biaisés et sous-estiment la variance de l'erreur. La puissance des tests d'hypothèses associés est alors surestimée. C'est pourquoi on leur préfère maintenant l'une des deux approches suivantes : par maximum de vraisemblance (*full information maximum likelihood*) ou par imputations multiples (*multiple imputation*) (Enders, 2010). Ce ne sont toutefois pas tous les logiciels qui permettent l'application de ces méthodes ou qui permettent leur application à toutes les méthodes d'analyse.

4.3. Linéarité des relations

Comme pour la régression multiple, la relation entre les variables dépendantes et indépendantes doit être linéaire. Puisque les variables indépendantes sont toutes catégorielles, la linéarité entre ces variables de même qu'entre celles-ci et les variables dépendantes ne pose jamais de problème. Toutefois dans le contexte de la MANOVA, il faut s'assurer de la linéarité des variables dépendantes entre elles. Cette vérification peut être effectuée de maintes façons. La plus simple consiste à inspecter un diagramme de dispersion qui présentera la relation entre les valeurs de chacune des variables dépendantes. Il est aussi possible de comparer le coefficient de corrélation de Pearson au coefficient de corrélation de Spearman, basé sur les rangs. En outre, on peut effectuer une régression non paramétrique qui permettra d'obtenir la forme de la relation la plus plausible entre les variables (voir la figure 16.2 dans l'exemple à la fin du chapitre). Si la relation entre les variables n'est pas linéaire, mais tout de même monotone (c'est-à-dire croissante ou décroissante), la valeur du coefficient de corrélation de Spearman sera supérieure à celle du coefficient de Pearson. Lorsque la relation ne peut pas être considérée comme linéaire, il peut être nécessaire de procéder à une transformation non linéaire des variables.

4.4. Indépendance des observations

Cette condition d'application de l'ANOVA et de la MANOVA implique qu'il n'y ait pas de lien entre les observations. Autrement dit, une observation ne devrait pas avoir d'influence sur le choix d'une autre ni sur les résultats de cette dernière. Pour favoriser l'indépendance des observations, il est possible de recourir à une méthode d'échantillonnage aléatoire[7] et de répartir au hasard les participants au sein de chacun des groupes. En dépit de toutes ces précautions, il se peut que les observations puissent s'influencer mutuellement notamment lorsque le traitement implique des interactions (p. ex., discussion) entre des personnes (Glass et Hopkins, 1984). Le lecteur qui souhaite approfondir cette question peut consulter l'article de Kenny et Judd (1986).

Selon Hair et ses collaborateurs (2010), aucun test statistique n'est en mesure de détecter l'ensemble des formes de dépendance possibles. Lorsque le chercheur a un doute à ce sujet, il devrait envisager une autre méthode d'analyse. C'est pourquoi l'analyse à niveaux multiples ou multiniveaux

7. L'échantillonnage devra être effectué avec remise ou à partir d'une population contenant suffisamment d'éléments pour que l'extraction de l'un d'eux n'altère pratiquement pas la probabilité de sélection des autres.

(Raudenbush et Bryk, 2002 ; Snijders et Bosker, 1999), qui tient compte de la dépendance entre les observations, remplace de plus en plus la MANOVA. De plus, elle n'est pas affectée par la présence de données manquantes qui, elle aussi, provoque une dépendance entre les observations.

4.5. Normalité de la distribution

Bien que le principe relatif à la décomposition de la variance à la base du test F demeure valable peu importe la distribution des variables étudiées, l'application de l'ANOVA et de la MANOVA repose également sur la normalité de la distribution. Plus spécifiquement, les résultats doivent provenir d'une population où ils se distribuent normalement. Ainsi, la distribution de la variable dépendante, ANOVA, ou des variables dépendantes, MANOVA, devrait respectivement être normale et multinormale pour chacune des catégories de facteurs.

Différentes approches permettent de vérifier le respect de ce postulat : des représentations graphiques (p. ex., histogramme, diagramme probabilité-probabilité), le calcul de coefficients d'asymétrie et d'aplatissement, les tests de Kolmogorov-Smirnov ou de Shapiro-Wilk, etc.

La façon la plus simple de vérifier la normalité est sans doute d'inspecter visuellement l'histogramme qui compare les données recueillies à une approximation de la distribution normale. Bien que cette méthode soit séduisante en raison de sa simplicité, elle est problématique dans le cas des petits échantillons, car le nombre de catégories pour construire l'histogramme ou la largeur de celles-ci peuvent en fausser la représentation visuelle (Hair *et al.*, 2010).

Une approche plus fiable est le diagramme probabilité-probabilité ou quantile-quantile. Il permet d'évaluer la concordance entre la distribution des données et une distribution théorique, dans le cas qui nous intéresse, la distribution normale. Dans ce graphique (voir la figure 16.3 dans l'exemple à la fin du chapitre), l'ordonnée représente les fréquences cumulées des données observées alors que l'abscisse porte sur les probabilités cumulées correspondant à la loi normale. Ainsi, la distribution normale forme une ligne droite en diagonale, et les données sont comparées à celle-ci. Les écarts de la distribution normale en termes de coefficient d'aplatissement sont facilement visibles dans les tracés de normalité. Lorsque les données tombent sous la diagonale, la distribution observée est plus plate que prévu. Quand elles passent au-dessus de celle-ci, la distribution est plus pointue que la courbe normale.

En plus des représentations graphiques, les coefficients d'asymétrie et d'aplatissement peuvent également servir à évaluer la normalité de la distribution. Ceux-ci sont habituellement disponibles dans la plupart des logiciels statistiques. Un coefficient de 0 reflète généralement la normalité de la distribution. Plus la valeur de ces coefficients s'éloigne de 0 (autant dans les valeurs positives que négatives), plus la distribution des données présente un écart par rapport à la normalité. À titre indicatif, une valeur d'asymétrie positive indique qu'il y a plus de résultats faibles dans la distribution que dans la distribution normale, alors qu'une valeur négative révèle, à l'inverse, davantage de résultats élevés que celle-ci. Une valeur d'aplatissement positive reflète une distribution plus pointue (appelée leptocurtique) que la distribution normale et une valeur négative révèle une distribution plus aplatie et étalée (nommée platicurtique) que celle-ci.

Ces coefficients peuvent être convertis en scores z en les divisant par leur erreur type (Kendall et Stuart, 1963). Ainsi, il devient possible de les comparer aux valeurs critiques d'une distribution normale. Les plus usuelles sont : ±2,58 (au seuil de signification de 0,01) et ±1,96 (au seuil de signification de 0,05). Si le résultat obtenu dépasse ces valeurs critiques en valeur absolue (c'est-à-dire sans en considérer le signe), la distribution des données peut alors être considérée comme significativement différente de la normalité, en termes d'asymétrie ou d'aplatissement, au seuil de probabilité correspondant. Or, ces tests de signification ne devraient pas être utilisés dans le cas de grands échantillons, car ils risquent d'être trop facilement significatifs, en raison de la puissance statistique (probabilité de détecter une différence significative lorsqu'il y en a une), et ce, même lorsque l'asymétrie ou l'aplatissement ne diffèrent pas beaucoup de la normalité.

Deux autres tests statistiques peuvent également servir à vérifier l'hypothèse de la normalité. Il s'agit des tests de Kolmogorov-Smirnov[8] et de Shapiro-Wilk. Ceux-ci comparent les données de l'échantillon à un ensemble de résultats se distribuant normalement de mêmes moyennes et écarts types. L'hypothèse nulle de ces tests postule la normalité de la distribution. Si le test est non significatif ($p > 0,05$), cela signifie que la distribution de l'échantillon n'est pas significativement différente d'une distribution normale (c'est-à-dire qu'elle est probablement normale). Cependant, le rejet de celle-ci au seuil de signification retenu reflète le non-respect de cette condition d'application.

8. Lilliefors (1967) a proposé une table pour tester la normalité en utilisant la statistique de Kolmogorov-Smirnov quand la moyenne et l'écart type de la population ne sont pas connus et sont estimés à partir de l'échantillon.

Dans de grands échantillons, ces tests statistiques peuvent aussi donner lieu à des résultats significatifs, même s'ils diffèrent très légèrement d'une distribution normale. Selon Hair et ses collaborateurs (2010), en raison de la puissance statistique, ces tests sont peu utiles dans le cas de petits (moins de 30 observations) ou de grands échantillons (plus de 1 000 observations). C'est pourquoi, en pratique, le chercheur est souvent contraint d'opter pour d'autres avenues comme le diagramme probabilité-probabilité et les valeurs des coefficients d'asymétrie et d'aplatissement.

La normalité univariée de la distribution peut donc facilement être testée et un certain nombre de mesures correctives sont disponibles dont les transformations qui seront abordées plus loin. Toutefois, il importe de souligner qu'heureusement, l'ANOVA n'est pas très sensible aux écarts à la normalité. Autrement dit, il s'agit d'une analyse considérée robuste à l'égard de la violation de ce postulat lorsque les autres conditions sont remplies.

Selon Glass, Peckham et Sanders (1972), l'asymétrie a seulement un petit effet (généralement de quelques centièmes) sur la probabilité associée au test ou la puissance statistique. De plus, même si les effets de l'aplatissement sont habituellement plus grands, ils tendent à demeurer négligeables. D'ailleurs, le théorème central limite stipule que la somme d'observations indépendantes s'approche d'une distribution normale au fur et à mesure que leur nombre augmente, et ce, indépendamment de leur distribution initiale. Ainsi, dans la plupart des cas et au fur et à mesure que la taille des échantillons augmente, le chercheur devient moins préoccupé par des variables qui ne se distribuent pas normalement, sauf si les autres postulats ne sont pas respectés (voir sous-section « Homogénéité des variances et des matrices de covariances »).

La MANOVA requiert aussi que les données suivent une distribution normale, mais multivariée. Celle-ci implique non seulement que la distribution de chacune des variables soit normale, mais que ce soit également le cas de toutes leurs combinaisons linéaires. Or, la distribution normale multivariée n'est pas facile à tester et peu de tests statistiques sont disponibles dans les logiciels à cet effet (Stevens, 2009). C'est pourquoi les chercheurs se limitent souvent à la vérification de la distribution normale univariée pour chacune des variables à l'étude (Hair *et al.*, 2010). Bien que cette procédure ne garantisse pas que la distribution multivariée soit normale, les écarts à celle-ci ont généralement peu de conséquences lorsque la normalité est respectée pour chacune des variables dépendantes (Hair *et al.*, 2010), surtout si la taille des groupes est la même et qu'elle est assez importante. Les lecteurs souhaitant approfondir cette question ou identifier des solutions de rechange peuvent consulter Stevens (1986) ou von Eye et Bogat (2004).

En somme, au-delà de la distribution de probabilités, il importe de tenir compte de la taille des groupes pour juger du non-respect de la condition de normalité.

Comme c'est le cas pour l'ANOVA, selon Hair et ses collaborateurs (2010), la violation de ce postulat a généralement peu d'effet dans le contexte de grands échantillons. De plus, la violation de celui-ci crée principalement des problèmes lors de l'application du test de Box. Cependant, des transformations peuvent corriger ce problème dans la plupart des situations. Lorsque la taille des échantillons est modérée, les écarts à la normalité peuvent être acceptables dans la mesure où ceux-ci sont liés à l'asymétrie et non pas à des données extrêmes ou influentes.

4.6. Homogénéité des variances et des matrices de covariances

Lors des calculs de l'ANOVA, la combinaison des variances évaluées séparément pour chacun des groupes sert d'estimé à la variance globale. Conséquemment, il importe que la condition d'homoscédasticité soit respectée, c'est-à-dire que la variance de la variable dépendante soit homogène à travers l'ensemble des groupes à comparer. Pour certains auteurs (Raudenbush et Bryk, 2002), c'est le non-respect de ce postulat qui peut le plus affecter la validité des résultats obtenus à une ANOVA.

Afin de vérifier ce postulat, il est possible de produire des représentations graphiques (histogramme et diagramme en boîte et moustaches) pour chacun des groupes ou d'appliquer un test statistique à cet effet. Contrairement aux tests de Bartlett et de Hartley, celui de Levene ne requiert pas que la distribution des échantillons soit normale (Glèlè Kakaï, Sodjinou et Fonton, 2006). L'hypothèse nulle du test de Levene stipule l'homogénéité des variances entre les groupes. Le rejet de celle-ci au seuil de signification retenu (souvent 0,05) reflète le non-respect du postulat.

Comme pour les tests de normalité, en raison de la puissance du test, lorsque la taille des groupes est grande, un petit écart entre les groupes peut donner lieu à une différence statistique. Lorsque c'est le cas, il peut être utile de calculer le F_{max} de Hartley (Hartley, 1950). Celui-ci compare la variance du groupe qui présente le plus de variabilité avec celle du groupe en ayant le moins. Après quoi, la valeur obtenue est comparée à des valeurs critiques dans une table des scores F.

Lorsque le postulat d'homoscédasticité n'est pas respecté, il est possible d'envisager une transformation des données (voir section « Transformation des données ») ou l'application de corrections, particulièrement lorsque la taille des groupes est très inégale. À cet égard, le logiciel SPSS donne

accès aux corrections de Welch et de Brown-Forsythe ainsi qu'à des tests de comparaisons multiples qui ne reposent pas sur cette condition d'application (T2 de Tamhane, T3 de Dunnett, Games-Howell ou C de Dunnett).

Pour ce qui est de la MANOVA, comme il y a plusieurs variables dépendantes, il importe de s'assurer que la variance des groupes est similaire, mais aussi que la covariance entre les variables dépendantes l'est également à travers les cellules du devis. Cela concerne donc maintenant l'homogénéité des matrices de covariances à travers les groupes.

Plusieurs tests statistiques peuvent servir à évaluer le respect de cette condition. Le test M de Box est généralement utilisé. Toutefois, celui-ci est sensible aux déviations du postulat lié à la normalité multivariée (Tabachnick et Fidell, 2013). C'est pourquoi il est préférable de vérifier la normalité des distributions de chacune des variables dépendantes avant d'appliquer celui-ci.

Au même titre que les autres tests statistiques, le test de Box peut mener au rejet de l'hypothèse nulle même si les matrices de covariances sont semblables dans le cas de grands échantillons. C'est d'autant plus vrai, car il est sensible aux écarts de normalité.

Contrairement à l'ANOVA, l'effet de la violation du postulat de l'homogénéité des matrices de covariances dans le contexte d'une MANOVA n'est pas clair sauf pour le T^2 de Hotelling lorsque la taille des groupes est égale (Hakstian, Roed et Lind, 1979). En extrapolant cette information, Tabachnick et Fidell (2013) font état de la robustesse de la MANOVA lorsque les échantillons sont égaux. Toutefois, selon eux, celle-ci n'est plus assurée si la taille des groupes diffère et que le test de Box s'avère significatif au seuil de probabilité de 0,001. D'après Hair et ses collaborateurs (2010), la violation de ce postulat entraîne peu de conséquences si les groupes sont de taille à peu près équivalente, c'est-à-dire si le rapport entre le groupe le plus grand et le plus petit est inférieur à 1,5. Tout comme pour la normalité, l'homogénéité des matrices de covariances peut être corrigée par le biais de transformations de données ou de corrections aux tests d'hypothèses appliqués.

5. TRANSFORMATION DES DONNÉES

La transformation des données s'avère souvent une solution efficace lorsque les conditions d'application de normalité de la distribution et d'homoscédasticité de l'ANOVA ou d'homogénéité des matrices de covariances avec la MANOVA ne sont pas remplies. En pratique, le chercheur doit identifier, parmi l'ensemble des transformations possibles, celle qui lui permet d'en

arriver à respecter les postulats requis. Autrement dit, ceux-ci doivent être revérifiés à partir des données transformées avant de procéder à l'ANOVA ou à la MANOVA.

Comme la violation des postulats est souvent liée à la distribution de la variable dépendante, Tabachnick et Fidell (2013) font des recommandations de transformations en fonction de la nature de l'asymétrie. Lorsque l'asymétrie est positive, ils proposent des transformations par la racine carrée, logarithmique et inverse (1/Y) selon son intensité, modérée, importante ou sévère. Ils suggèrent aussi de recourir à ces transformations pour l'asymétrie négative, mais cette fois-ci, après avoir inversé les résultats à l'échelle en appliquant le reflet comme transformation. Pour ce faire, il suffit de prendre la plus grande valeur de la distribution et de lui soustraire le résultat obtenu à la variable dépendante (K – Y). À la suite de ce calcul, il importe d'ajouter la constante 1 pour éviter que le résultat minimal soit égal à 0, car dans ce cas, il devient impossible d'appliquer certaines transformations (p. ex., logarithmique, inverse). Notons qu'une transformation des données peut inverser l'interprétation des résultats. Une façon de s'assurer qu'aucune erreur n'est commise à cet égard consiste à interpréter ceux-ci à partir des données brutes ou non transformées.

Hair et ses collaborateurs (2010) soulignent aussi qu'en pratique certaines transformations ont été associées à des types de données en particulier. Par exemple, une transformation par la racine carrée est souvent appliquée dans le cas de fréquences (p. ex., nombre d'erreurs, nombre de réponses) alors qu'une transformation angulaire (arcsin) peut s'avérer particulièrement efficace dans le cas de proportions. De même, les unités de temps (p. ex., minutes, secondes) et les changements de proportions peuvent être transformés par le biais du logarithme de la variable.

Après avoir fait de multiples tentatives, il peut arriver que les transformations des données ne permettent pas de respecter les postulats à l'étude. Dans le cas des ANOVA, des auteurs (Glass *et al.*, 1972) suggèrent de recourir aux tests non paramétriques (Kruskal-Wallis ou de Friedman). Or, ces tests ne poursuivent pas exactement le même objectif et ne peuvent pas être considérés comme des tests équivalents à une ANOVA au sens strict. De même, bien que leur présentation aille au-delà des objectifs poursuivis par ce chapitre, des tests statistiques sont disponibles lorsque les conditions de normalité multivariée et d'homogénéité des matrices de covariances ne peuvent être assumées (Zwick, 1985).

Il faut toutefois souligner que l'utilisation de transformations dans le contexte d'une MANOVA peut avoir un impact important sur la linéarité entre les variables. Par exemple, il peut arriver qu'on corrige un problème

avec une transformation *arcsin*, mais que d'un autre côté, cela fasse en sorte que la relation entre certaines variables ne soit plus linéaire. Il faut donc être très prudent avec l'utilisation des transformations. De plus, il est généralement impossible d'interpréter les variables transformées et il faut alors les ramener à leur échelle d'origine avant de procéder à cette interprétation.

6. DÉMARCHE D'ANALYSE UNIVARIÉE

L'ANOVA est une technique d'analyse qui repose sur un processus en deux étapes (Bray et Maxwell, 1982). La première consiste à tester l'hypothèse nulle stipulant qu'il n'y a pas de différence dans les moyennes entre les groupes. Si le résultat à ce test s'avère significatif, la deuxième étape vise à effectuer d'autres tests pour expliquer les différences entre ceux-ci. Autrement dit, l'ANOVA est fort utile pour vérifier si plusieurs moyennes sont différentes les unes des autres.

6.1. Procédures de comparaisons multiples

Dans le cas où l'hypothèse nulle est rejetée, il est possible de conclure que les moyennes des groupes ne sont pas toutes identiques. Or, ces analyses ne permettent pas d'identifier quelles moyennes sont différentes les unes des autres. S'il n'y a que deux groupes, il suffit de consulter les moyennes de chacun d'eux pour savoir lequel obtient une moyenne supérieure à l'autre. Toutefois, lorsqu'il y a plus de deux groupes, il peut arriver que seulement certains d'entre eux soient statistiquement différents des autres. Afin d'identifier entre quels groupes se situent la ou les différences, il faut recourir à des comparaisons multiples *a posteriori* (appelées aussi *post-hoc*).

6.1.1. Comparaisons *a posteriori*

On peut se demander ici pourquoi on ne fait pas une série de tests *t* pour évaluer les différences entre les groupes deux à deux. Cela s'explique par le fait qu'en réalisant plusieurs tests de comparaison sur les mêmes moyennes, la probabilité de trouver un résultat significatif par pur hasard augmente, c'est ce qu'on appelle l'erreur d'ensemble, erreur qui consiste en une accumulation des erreurs par test (seuil de signification statistique usuel). Pour remédier à ce problème, le niveau de signification est généralement corrigé en fonction du nombre de comparaisons à effectuer, d'où l'intérêt des procédures de comparaisons multiples.

La correction la plus souvent appliquée est celle de Bonferroni (Neter, Wasserman et Kutner, 1985), car il s'agit de la plus simple. En effet, il suffit de diviser le seuil de signification (habituellement 0,05) par le nombre total de comparaisons à réaliser.

Plusieurs autres tests de comparaisons multiples sont utilisés. Ils diffèrent essentiellement dans leur façon de contrôler l'erreur de type I. En ce sens, certains sont plus permissifs ou, au contraire, plus conservateurs que les autres. Or, le niveau d'erreur de type I et la puissance sont interreliés. En réduisant la probabilité d'erreur de type I (conclure qu'il y a une différence quand il n'y en a pas dans la réalité), la puissance du test s'en trouve limitée (probabilité de détecter une différence statistiquement significative). Le chercheur doit donc faire un compromis.

Du plus conservateur au plus permissif, les tests de comparaisons multiples les plus fréquents sont le test de Scheffé, de la différence franchement significative de Tukey HSD (*honestly significant difference*), de la plus petite différence significative de Fisher LSD (*least significant difference*), de Newman-Keuls et de Dunnett (Howell, 2008). Toutes ces procédures permettent de contrôler l'erreur d'ensemble en diminuant le nombre de comparaisons multiples et en ne conservant que les plus utiles. Par exemple, avec le test de Tukey, en premier lieu, on place en rang les différences de moyennes. Ensuite, on effectue un test d'hypothèses à partir de la différence la plus importante. On arrête la procédure dès qu'on ne peut pas rejeter l'hypothèse nulle à l'une des différences dont le rang est plus faible. De cette façon, seulement un sous-ensemble des comparaisons est utilisé. Le test de Newman-Keuls est une variante du test de Tukey où le seuil de signification associé à chaque erreur par test diminue lorsque le rang de la différence de moyennes augmente. Ainsi, puisque la première différence testée est la plus importante et est statistiquement significative, la deuxième différence de moyennes sera moins importante et le seuil de signification utilisé pourra être plus petit. Pour sa part, le test de Scheffé réduit le nombre de comparaisons multiples en utilisant des contrastes plus complexes qui vont au-delà de la simple comparaison de deux moyennes. Le test de Dunnett est particulièrement intéressant lorsqu'il y a un groupe témoin, car il limite le nombre de comparaisons en ne confrontant chacune des moyennes qu'avec celle issue de celui-ci.

6.1.2. Comparaisons *a priori*

Les comparaisons *a posteriori* sont fréquemment utilisées, car elles sont faciles à appliquer. Toutefois, s'il est préalablement possible d'émettre des hypothèses spécifiques à l'égard des différences attendues entre les groupes,

des comparaisons *a priori* devraient être privilégiées. Contrairement aux comparaisons *a posteriori*, qui sont effectuées une fois que le test F global s'est avéré statistiquement significatif, les comparaisons *a priori*, comme leur nom l'indique, se définissent en amont de l'analyse sur la base d'un modèle théorique. En fait, elles se substituent généralement au test F global.

La façon de les calculer ressemble à celle des comparaisons *a posteriori*, mais s'en distingue à plusieurs égards. Le chercheur doit avoir une idée préalable des comparaisons de groupes qu'il souhaite effectuer. En fait, il ne devrait pas être intéressé par l'ensemble des comparaisons possibles, comme c'est le cas pour les comparaisons *a posteriori*. Pour cette raison, les tests d'hypothèses associés à ces comparaisons sont plus puissants, car les corrections appliquées au seuil de signification par les tests permettent de tenir compte d'un nombre généralement fort restreint de comparaisons réalisées. Toutefois, les comparaisons *a priori* ne devraient être utilisées que lorsqu'elles reposent sur des bases théoriques.

Il existe plusieurs méthodes de comparaisons multiples *a priori*. La plus simple est sans doute d'utiliser des tests *t* et d'appliquer la correction de Bonferroni présentée ci-dessus.

La présentation détaillée des tests de comparaisons multiples (*a priori* ou *a posteriori*) disponibles dépasse largement les objectifs de ce chapitre. Le lecteur intéressé par cette question peut consulter les écrits de Howell (2008), Huitema (1980) ou Toothaker (1993) pour obtenir de plus amples informations.

6.2. Effets d'interaction et effets simples

Comme il en a été question précédemment, il arrive que des plans factoriels d'ANOVA et de MANOVA soient utilisés pour vérifier l'effet de plusieurs variables indépendantes ainsi que de ceux liés à leur interaction. Lorsque l'effet d'une seule variable est considéré à la fois, il est question de son effet principal, alors que lorsque l'effet de plusieurs variables est pris en compte simultanément, il s'agit d'effets d'interaction.

Un effet d'interaction survient ou s'avère significatif lorsque l'effet d'une variable indépendante sur une variable dépendante varie en fonction des niveaux d'une autre variable indépendante. Autrement dit, une interaction est présente quand le patron des résultats pour un facteur est différent pour les divers niveaux de l'autre facteur.

Afin de comprendre la nature de l'interaction entre deux variables indépendantes, il est généralement d'usage de recourir à une représentation graphique. La caractéristique clé des effets d'interaction réside dans le fait que les droites ne sont pas parallèles. De plus, afin de vérifier si l'effet de l'une des variables indépendantes est significatif pour chacun des niveaux de l'autre (pour un niveau fixe de ce facteur), l'analyse des effets simples est habituellement réalisée. Ce type d'analyse permet d'approfondir ce qui se passe dans chacune des conditions. Or, en pratique, il faut être prudent, car plus nous testons d'effets simples, plus le risque de commettre une erreur de type I s'accroît. Ainsi, les chercheurs choisissent de maintenir constante l'une ou l'autre des variables indépendantes aux fins de ces analyses.

Quand l'effet d'interaction est significatif, les effets principaux de chacune des variables indépendantes ne sont généralement pas interprétés à moins qu'ils soient vraiment pertinents pour la recherche en cours, car leur interprétation peut ne pas avoir beaucoup de sens en présence d'un effet d'interaction. Lorsque l'effet d'interaction n'est pas significatif, les droites sont considérées comme parallèles. Toutefois, il est possible d'observer des effets principaux significatifs ou des différences en fonction de chacune des variables indépendantes.

6.3. Taille des effets et puissance statistique

Suivant un résultat significatif à l'ANOVA et à la MANOVA, qu'il s'agisse d'un effet principal, d'un effet d'interaction ou d'un effet simple, il importe d'en évaluer l'importance. Pour ce faire, l'indice statistique le plus utilisé est le coefficient êta^2 partiel (η^2). Il constitue une mesure de ce qu'on appelle la taille de l'effet. Quand cet indice est appliqué à chacun des effets ou aux effets d'interaction, on les appelle généralement des indices partiels de la taille de l'effet. Lorsqu'il est appliqué à l'ensemble de l'analyse, on ne fait pas cette distinction.

Le coefficient êta^2 est strictement équivalent au coefficient de détermination R^2 et sert à évaluer la proportion de variance de la ou des variables dépendantes expliquée par un effet. Il consiste à calculer le rapport entre la variabilité intergroupes et la variabilité totale. Ainsi, sa valeur peut osciller de 0 à 1. Selon l'interprétation du R^2 de Cohen (1988), des valeurs de 0,01, de 0,09 et de 0,25 reflètent respectivement de petits, de moyens et de grands effets.

Enfin, il importe de noter que la probabilité d'obtenir un résultat statistiquement significatif dans des conditions similaires correspond à la puissance du test. Puisque plusieurs tests d'hypothèses sont appliqués dans

une ANOVA ou une MANOVA, la puissance peut être calculée pour chacun des effets ou encore des comparaisons *a posteriori*. De plus, puisqu'il s'agit d'une probabilité, la valeur de la puissance peut varier de 0 à 1. Elle est plus près de 0 lorsque la probabilité d'obtenir un résultat statistiquement significatif dans des conditions similaires est faible et plus près de 1 dans le cas inverse. On peut accroître la puissance des tests en augmentant la valeur du seuil de signification ainsi que de la taille de l'échantillon. On espère généralement obtenir une puissance d'au moins 0,80.

7. DÉMARCHE D'ANALYSE MULTIVARIÉE

Comme l'ANOVA, la MANOVA est une technique d'analyse qui repose sur un processus en deux étapes (Bray et Maxwell, 1982). Le test F permet seulement de déterminer si globalement, il y a une différence entre les vecteurs de moyennes, c'est-à-dire si l'hypothèse nulle est rejetée ou non. Si le résultat s'avère significatif, la deuxième étape consiste à effectuer d'autres tests pour expliquer les différences. En effet, il devient intéressant d'identifier quelles variables dépendantes participent à cette différence entre les groupes, car ce n'est probablement pas le cas de l'ensemble de celles-ci. Or, il n'y a pas de consensus dans les écrits sur la meilleure façon de procéder à ces analyses dans le contexte d'une MANOVA. À cet effet, deux avenues principales sont généralement empruntées par les chercheurs : la réalisation de multiples ANOVA ou d'une analyse discriminante.

7.1. Multiples ANOVA

L'une des procédures les plus fréquemment utilisées consiste à poursuivre l'analyse des données par le biais d'ANOVA sur chacune des variables dépendantes. On peut alors se demander pourquoi il importe de faire une MANOVA plutôt que de procéder directement à de multiples ANOVA. À ce sujet, Bock (1975) mentionne que cela permet au chercheur d'exercer un certain niveau de contrôle sur l'erreur de type I. En effet, si l'analyse globale (MANOVA) n'est pas significative, il ignorera les résultats des tests univariés (ANOVA).

Or, la MANOVA reflète uniquement un effet significatif sur la combinaison linéaire de variables dépendantes. Cela ne signifie pas pour autant que chacune de celles-ci soit significative. C'est pourquoi certains chercheurs (p. ex., Harris, 1975) suggèrent d'appliquer la correction de Bonferroni pour réajuster le seuil de signification avec lequel les résultats des ANOVA seront évalués.

Enfin, il importe de noter que lorsque des comparaisons multiples sont réalisées à la suite d'ANOVA, la probabilité d'erreur de type I s'accroît encore. Cela peut éventuellement poser de sérieux problèmes lorsque les variables dépendantes sont corrélées. De plus, le lecteur doit garder à l'esprit que dans le cas où les contributions des variables se recoupent, les statistiques univariées ne peuvent servir à départager leur contribution respective.

7.2. Analyse discriminante

Une procédure, plus conforme à la philosophie sous-jacente à la MANOVA, consiste à réaliser une analyse discriminante après celle-ci pour en interpréter les résultats. Contrairement aux calculs de multiples ANOVA, cette procédure offre l'avantage de tenir compte des relations qui existent entre les variables dépendantes. Plus spécifiquement, elle sert à identifier le nombre de combinaisons linéaires de variables, ou fonctions discriminantes, qui permettent de différencier significativement les groupes. Une fois ces combinaisons identifiées, il devient possible de cerner la contribution des variables dépendantes à chacune d'elles et de préciser les groupes qu'elles permettent de distinguer. Cette approche ne sera pas approfondie dans le présent chapitre pour éviter la redondance avec celui traitant de l'analyse discriminante. Le lecteur intéressé par cette approche d'analyse pourra consulter le chapitre 17 de cet ouvrage.

8. ILLUSTRATION DE L'ANALYSE DES PROBLÈMES DE COMPORTEMENT CHEZ LES JEUNES AYANT UNE DÉFICIENCE INTELLECTUELLE SELON LE SEXE ET LE NIVEAU DE DÉFICIENCE

Les étapes d'analyse et l'interprétation d'une MANOVA seront illustrées à partir d'un exemple pratique. Cette démarche permettra également de démontrer les étapes reliées au calcul de l'ANOVA.

Imaginons qu'un chercheur souhaite vérifier si les problèmes de comportement des jeunes ayant une déficience intellectuelle varient en fonction de leur sexe (féminin ou masculin) et de leur niveau de déficience (légère, moyenne, grave ou profonde). Aux fins de son étude, il devra choisir la façon dont seront mesurés les problèmes de comportement. Or, ceux-ci ne s'évaluent pas aussi simplement que l'âge ou le poids, car ce concept n'est pas unidimensionnel. Il décide alors de demander à des mères de jeunes d'âge scolaire recevant des services d'un centre de réadaptation en déficience intellectuelle et en troubles envahissants du développement de remplir la version française du *Developmental Behavior Checklist* (DBC ; Einfeld et

Tonge, 2002). Il s'agit d'un instrument constitué de 96 items ayant pour but de dépister les problèmes de comportement chez les jeunes ayant une déficience intellectuelle âgés de 4 à 18 ans. Une échelle en 3 points est utilisée (0 = Pas vrai ou ne s'applique pas, 1 = Assez vrai ou parfois vrai, 2 = Très vrai ou souvent vrai) pour l'évaluation de chacun des comportements. Les réponses aux items sont ensuite additionnées et permettent de dégager des résultats en fonction de cinq types de problèmes de comportement : perturbateurs/antisociaux (DISRUP), centrés sur soi (SELFAB), troubles de la communication (COMPDIS), anxiété (ANXIET) et problèmes sociaux (SOCIAL). Ceux-ci constituent donc les variables dépendantes à l'étude.

Si le chercheur réalise cette étude, c'est parce qu'il croit que les problèmes de comportement des jeunes peuvent différer significativement en fonction de leur sexe et de leur niveau de déficience intellectuelle. Il soupçonne même que ces deux variables peuvent interagir. Autrement dit, il est possible que les problèmes présentés par les filles ou les garçons diffèrent, mais que cette différence dépende aussi du niveau de déficience intellectuelle. Les différences entre eux pourraient aussi être attribuables à leur âge, mais la taille de l'échantillon ne permettait pas de considérer cette variable additionnelle. Or, il aurait pu en contrôler l'effet comme covariable. Toutefois, cette avenue n'a pas été envisagée dans le présent exemple, car elle dépasse l'objectif poursuivi dans ce chapitre.

S'il désire mettre à l'épreuve ses hypothèses au moyen d'une analyse univariée, il pourrait effectuer cinq ANOVA séparées. Bien que cette démarche lui permette de vérifier ses hypothèses, elle présente des inconvénients importants. D'abord, les ANOVA réalisées sur chacune des variables ne tiennent pas compte de la corrélation entre les variables dépendantes (cinq types de comportements). Conséquemment, ces analyses risquent d'être redondantes ou de ne pas être aussi indépendantes qu'il l'espérait. Le chercheur aurait aussi pu réaliser une seule ANOVA sur le score global de l'échelle, mais il n'aurait alors pas pu faire la distinction entre les différents types de problèmes de comportement. Comme les problèmes de comportement présentent plusieurs dimensions, il est préférable de conserver l'ensemble des variables plutôt que d'en retenir une seule. De plus, en réalisant une MANOVA, qui tient compte de la corrélation entre les variables dépendantes, il se peut qu'il détecte globalement une différence significative entre les groupes qu'il n'aurait pas pu déceler par des ANOVA effectuées isolément sur l'une ou l'autre des variables dépendantes.

8.1. Vérification des conditions d'application

Préalablement à l'analyse, il doit vérifier si les données qu'il a recueillies respectent les diverses conditions d'application requises par la MANOVA. D'abord, il est d'usage de s'assurer qu'il y a suffisamment de participants dans chacun des groupes avant de procéder. En examinant le tableau 16.1, on remarque que le nombre de filles ayant une déficience intellectuelle profonde est limité (n = 7). C'est pourquoi le chercheur a pris la décision de regrouper les catégories « grave » et « profonde » de la variable *Niveau de déficience intellectuelle* aux fins des analyses. De plus, comme il ne s'agit pas d'un devis expérimental, le nombre de jeunes n'est pas le même dans chacune des cellules du tableau. Autrement dit, les groupes ne sont pas équivalents en nombre.

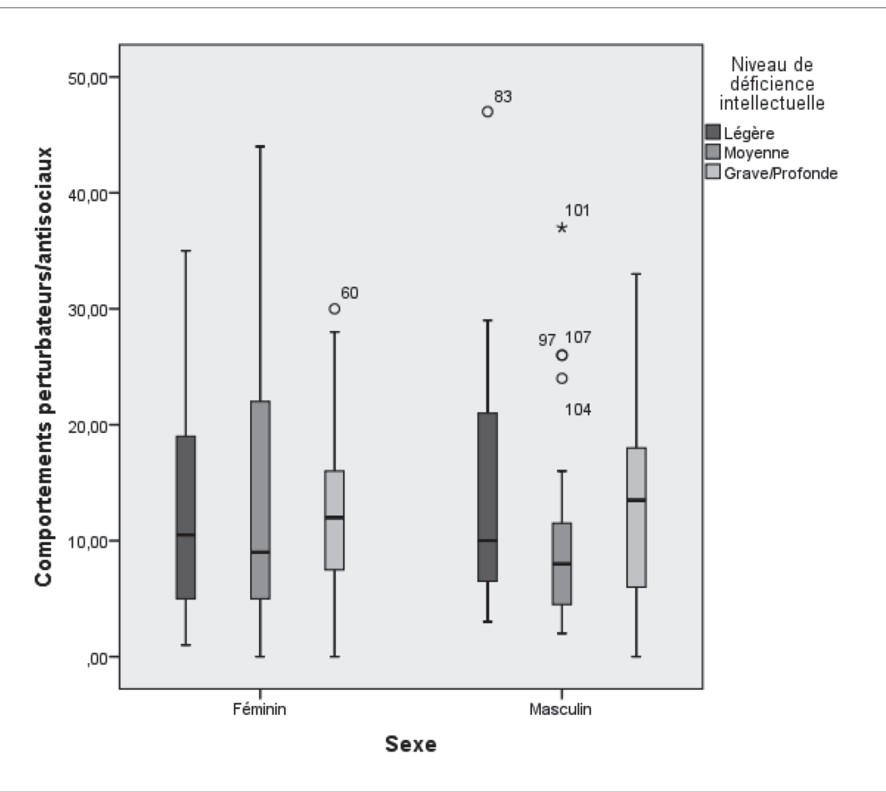
Tableau 16.1.
Nombre de jeunes en fonction du sexe et du niveau de déficience intellectuelle

	Niveau de déficience intellectuelle				Total
	Légère	**Moyenne**	**Grave**	**Profonde**	
Sexe Féminin	14	32	13	7	66
Masculin	19	23	20	14	76
Total	33	55	33	21	142

Dans un deuxième temps, le chercheur procède à l'exploration des données dans le but de déterminer la présence de valeurs manquantes et d'examiner la forme de la distribution ainsi que la variabilité dans les données pour l'ensemble des participants et chacun des groupes. À cette fin, il utilise des représentations graphiques, des tableaux de fréquences et des statistiques descriptives.

Par la suite, il procède à l'examen des valeurs extrêmes ou influentes. D'une part, le calcul des scores z l'a mené à identifier plusieurs valeurs extrêmes ou influentes univariées, c'est-à-dire supérieures à |3,00|. À titre d'exemple, la figure 16.1 illustre les résultats obtenus pour la variable comportements perturbateurs/antisociaux à l'aide d'un diagramme en boîte et moustaches. Chacun des points avec un numéro correspond à une valeur extrême ou influente.

Figure 16.1.
Diagramme en boîte et moustaches



D'autre part, la procédure Régression de SPSS l'a aussi conduit à repérer la présence de quelques valeurs extrêmes ou influentes multivariées. Dans le présent exemple, toute valeur de distance de Mahalanobis supérieure à 20,515 (c'est-à-dire à la valeur critique du χ^2, ayant 5 degrés de liberté au seuil de 0,001) peut être considérée comme atypique. C'est aussi le cas pour les valeurs de force de levier supérieure à 0,127, soit 3*(5 + 1)/142. Enfin, toute valeur de distance de Cook supérieure à 1 doit également être scrutée attentivement. Puisque ces conditions d'application ne sont pas spécifiques à la MANOVA, celles-ci ne sont pas illustrées dans le présent exemple. Le lecteur intéressé par plus de détails pourra consulter le livre de Tabachnick et Fidell (2013). Bien que plusieurs valeurs extrêmes ou influentes aient pu être identifiées au sein de l'échantillon, celles-ci ont été conservées aux fins des analyses, car elles reflètent la réalité de la population à l'étude (p. ex., après vérification de la saisie des données et du questionnaire rempli par la mère,

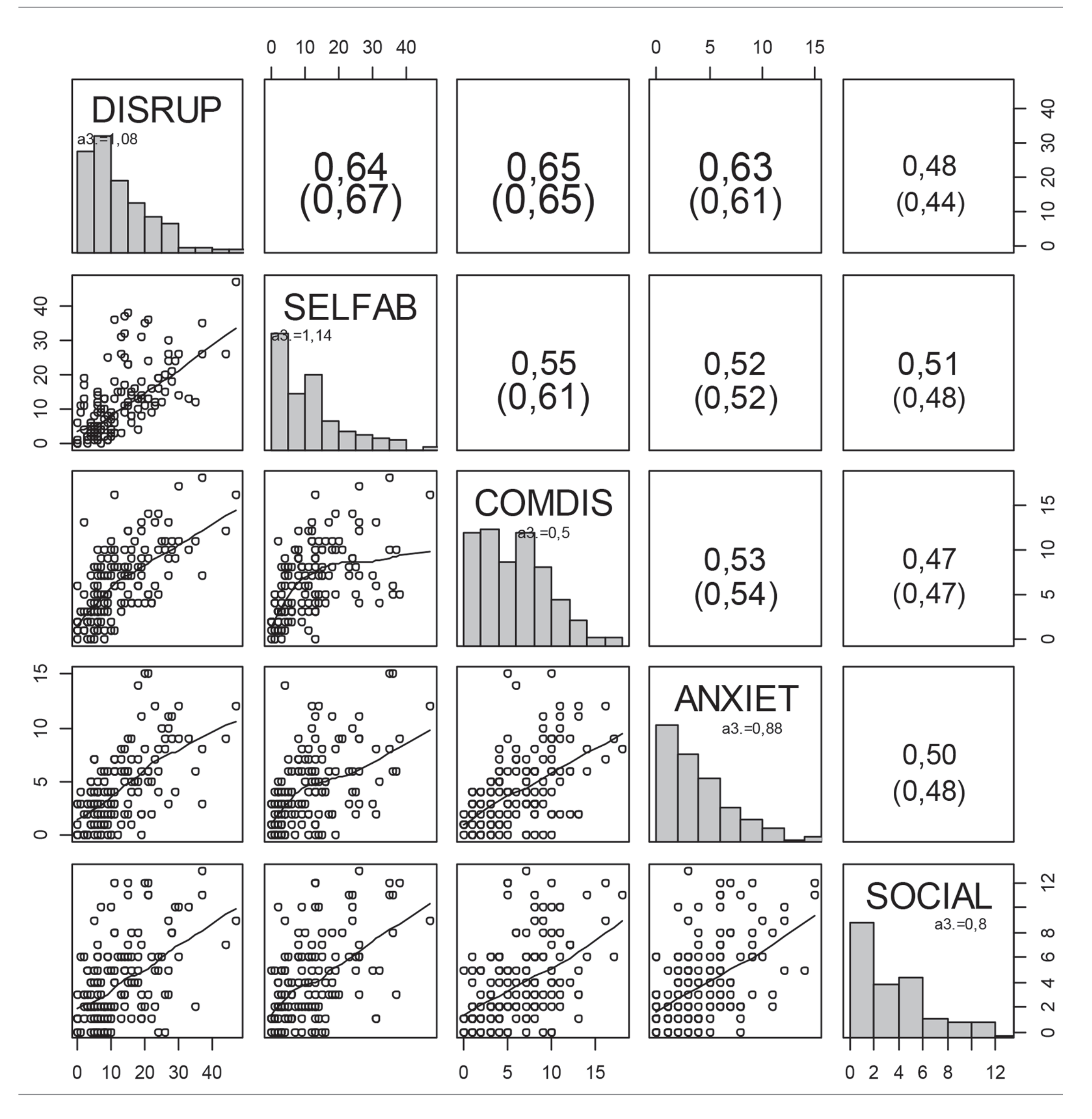
même si c'est peu typique pour une fille ayant une déficience intellectuelle grave ou profonde de présenter de nombreux comportements perturbateurs, cela demeure une possibilité). Dans la pratique, il faudrait examiner en détail le patron de réponses pour ces participants, éventuellement les retrancher de l'analyse, puis vérifier si les résultats obtenus convergent avec ceux des analyses incluant ces valeurs.

Comme le chercheur n'a pas vraiment de raison de croire que les données qu'il a recueillies ne sont pas indépendantes, il doit maintenant évaluer si les cinq indicateurs des problèmes de comportement entretiennent une relation linéaire entre eux et s'ils se distribuent normalement dans chacun des groupes. Dans le logiciel SPSS, la vérification de ces postulats doit être réalisée par d'autres commandes que celle visant le calcul de la MANOVA.

La figure 16.2 expose les résultats des diverses approches retenues pour vérifier si les cinq indicateurs des problèmes de comportement entretiennent une relation linéaire[9]. La diagonale illustre l'histogramme de fréquences de chacune des variables quantitatives ainsi que le coefficient d'asymétrie (a^3). On notera que toutes les variables affichent une asymétrie positive et donc à droite. À la diagonale supérieure sont présentés les coefficients de corrélation de Pearson et de Spearman (entre parenthèses). La différence entre ces coefficients est peu importante (au plus 0,06) et ce n'est pas sur cette base qu'on peut penser que la linéarité entre les variables pose problème. À la diagonale inférieure sont présentés les diagrammes de dispersion. La ligne qui passe à travers ces points représente la courbe de régression non paramétrique. La relation entre toutes ces variables est donc positive. Seul un diagramme de dispersion pourrait laisser supposer une relation sérieusement non linéaire. C'est celui du croisement entre les comportements centrés sur soi (SELFAB) et les troubles de communication (COMDIS). Selon la diagonale supérieure, c'est d'ailleurs à ce moment que la différence entre les coefficients de corrélation de Pearson et de Spearman est la plus grande.

9. Cette figure a été préparée à l'aide du logiciel R, car ce dernier permet de présenter l'ensemble des informations requises pour la vérification de ce postulat de façon très condensée.

Figure 16.2.
Analyse de la linéarité des relations entre les variables dépendantes



Pour ce qui est du postulat lié à la normalité des distributions de probabilités dans chacun des groupes, les résultats issus des tests de Kolmogorov-Smirnov et de Shapiro-Wilk révèlent que ce postulat n'est pas respecté dans plusieurs situations (tableau 16.2). En effet, plusieurs tests présentent une probabilité inférieure à 0,05, ce qui mène au rejet de l'hypothèse nulle

Tableau 16.2.
Tests de normalité pour chacun des types de problèmes de comportement en fonction du sexe et du niveau de déficience intellectuelle de l'enfant

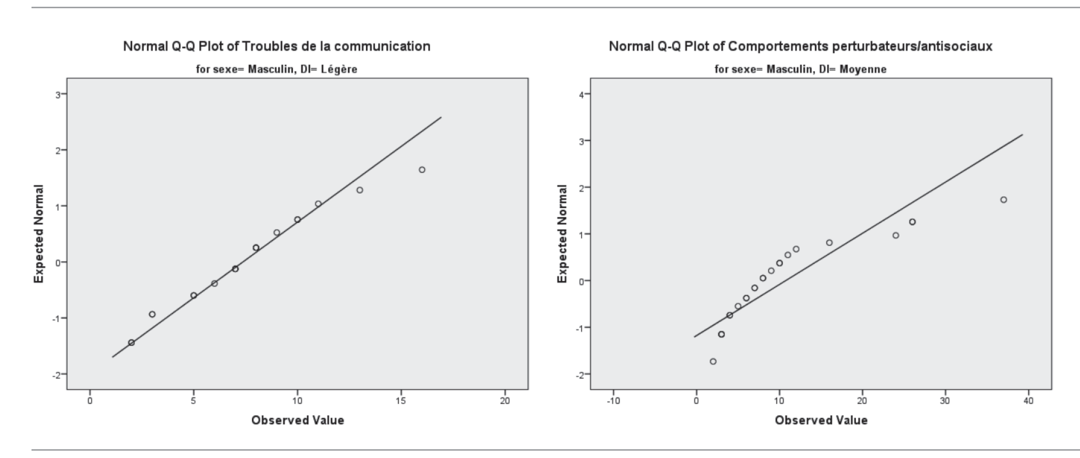
	Sexe	Niveau de déficience intellectuelle	Kolmogorov-Smimov[a]			Shapiro-Wilk		
			Statistic	df	Sig.	Statistic	df	Sig.
Comportements perturbateurs/ antisociaux	Féminin	Légère	0,183	14	0,200*	0,909	14	0,154
		Moyenne	0,185	32	0,007	0,903	32	0,007
		Grave/profonde	0,125	20	0,200*	0,951	20	0,390
	Masculin	Légère	0,202	19	0,039	0,849	19	0,006
		Moyenne	0,228	23	0,003	0,800	23	0,000
		Grave/profonde	0,087	34	0,200*	0,953	34	0,151
Comportements centrés sur soi	Féminin	Légère	0,191	14	0,179	0,886	14	0,071
		Moyenne	0,174	32	0,015	0,870	32	0,001
		Grave/profonde	0,111	20	0,200*	0,938	20	0,221
	Masculin	Légère	0,226	19	0,012	0,834	19	0,004
		Moyenne	0,219	23	0,005	0,854	23	0,003
		Grave/profonde	0,225	34	0,000	0,905	34	0,006
Troubles de la communication	Féminin	Légère	0,164	14	0,200*	0,892	14	0,087
		Moyenne	0,096	32	0,200*	0,952	32	0,166
		Grave/profonde	0,156	20	0,200*	0,935	20	0,190
	Masculin	Légère	0,116	19	0,200*	0,960	19	0,578
		Moyenne	0,174	23	0,071	0,916	23	0,055
		Grave/profonde	0,149	34	0,053	0,950	34	0,124
Anxiété	Féminin	Légère	0,196	14	0,148	0,910	14	0,157
		Moyenne	0,144	32	0,091	0,942	32	0,084
		Grave/profonde	0,159	20	0,199	0,935	20	0,192
	Masculin	Légère	0,264	19	0,001	0,778	19	0,001
		Moyenne	0,183	23	0,044	0,850	23	0,003
		Grave/profonde	0,183	34	0,005	0,898	34	0,004
Problèmes sociaux	Féminin	Légère	0,252	14	0,016	0,766	14	0,002
		Moyenne	0,243	32	0,000	0,881	32	0,002
		Grave/profonde	0,159	20	0,200*	0,937	20	0,214
	Masculin	Légère	0,142	19	0,200*	0,940	19	0,268
		Moyenne	0,208	23	0,011	0,833	23	0,001
		Grave/profonde	0,143	34	0,076	0,906	34	0,007

* This is a lower bound of the true significance.
a : Lilliefors Significance Correction.

stipulant que les données se distribuent normalement. De plus, il est à noter que les conclusions statistiques résultant de ces deux tests ne convergent pas toujours. Dans cet exemple, c'est le cas pour les problèmes sociaux chez les garçons ayant un niveau de déficience intellectuelle grave/profonde. En effet, en regardant les résultats au test de Kolomogorov-Smirnov, on peut constater que la probabilité est supérieure à 0,05 [D = 0,143, p = 0,076]. Conséquemment, on ne peut rejeter l'hypothèse nulle stipulant la normalité de la distribution. Or, le test de Shapiro-Wilk, pour le même groupe de garçons, nous amène à rejeter cette hypothèse, car la probabilité qui lui est associée est inférieure à 0,05 [W = 0,906, p = 0,007]. Autrement dit, selon ce test, la distribution ne peut pas être considérée comme normale. Il est à souligner que la sortie SPSS intitule la colonne où sont présentées les probabilités par *Sig*. C'est une erreur des concepteurs du logiciel, car bien sûr, il ne s'agit pas du niveau de signification qui doit être fixe et déterminé à l'avance. Cette erreur se retrouve dans la plupart des sorties informatiques obtenues à l'aide du logiciel SPSS.

La figure 16.3 présente deux diagrammes probabilité-probabilité. La figure de gauche illustre une variable dont la distribution s'apparente à une loi normale alors que la figure de droite montre une variable qui ne se distribue pas selon une loi normale.

Figure 16.3.
Diagramme comparatif probabilité-probabilité pour deux variables : l'une se comportant selon une loi normale, l'autre définitivement non



Pour ce qui est des postulats de l'homogénéité de la variance et de celle des matrices de covariances, les résultats aux tests de Box et de Levene sont disponibles lors du calcul de la MANOVA dans le logiciel SPSS. Comme la vérification des postulats est préalable à l'analyse, les résultats à ces tests sont présentés dès maintenant.

Rappelons d'abord que le test M de Box a pour but de vérifier l'hypothèse de l'homogénéité des matrices de covariances. Selon le tableau 16.3, le résultat à ce test est significatif ($p < 0{,}000$). Le chercheur ne peut donc pas conclure que les matrices de covariances sont homogènes. Conséquemment, le postulat de l'homogénéité des matrices de covariances n'est pas respecté. Or, il ne faut pas oublier que le test *M* de Box est sensible aux écarts de normalité et que, dans le présent exemple, la distribution de plusieurs groupes n'était pas considérée comme normale.

Tableau 16.3.
Résultats au test M de Box

Box's Test of Equality of Covariance Matrices[a]	
Box's M	172,677
F	2,059
df1	75
df2	17 617,252
Sig.	,000

Tests the null hypothesis that the observed covariance matrices of the dependant variables are equal across groups.
a : Design ; Intercept + sexe + Di + sexe * DI.

Un tableau récapitulatif des résultats au test d'égalité des variances de Levene pour chacune des variables dépendantes est également disponible lors du calcul de la MANOVA (tableau 16.4). Il importe de noter que les résultats qui s'y retrouvent sont les mêmes que ceux qui auraient été obtenus par des ANOVA sur chacune des variables dépendantes dans la mesure où il n'y a pas de valeur manquante parmi les données. S'il y avait eu des données manquantes, les analyses auraient dû être effectuées uniquement pour les observations où il n'y a pas de donnée manquante à aucune des variables dépendantes et indépendantes. En effet, le nombre d'observations peut être différent pour chacune des variables dépendantes ainsi que pour l'ensemble des variables dépendantes.

Tableau 16.4.
Résultats aux tests de Levene pour chacun des types de problèmes de comportement

	F	df1	df2	Sig.
Comportements perturbateurs/ antisociaux	1,850	5	136	,107
Comportements centrés sur soi	2,187	5	136	,059
Troubles de la communication	,769	5	136	,574
Anxiété	,538	5	136	,747
Problèmes sociaux	1,086	5	136	,371

Tests the null hypothesis that the observed covariance matrices of the dependant variables are equal across groups.
a : Design ; Intercept + sexe + Di + sexe * DI.

Dans cet exemple, le test de Levene est non significatif pour toutes les variables dépendantes examinées, car l'ensemble des probabilités calculées (ici Sig.) est supérieur à 0,05. Conséquemment, le postulat d'homogénéité des variances peut être considéré comme respecté.

8.2. Calcul de la MANOVA

Le tableau 16.5 montre le plan factoriel de l'analyse. On voit qu'il s'agit d'un plan 2 × 3 à 6 groupes indépendants (aucune mesure répétée). De plus, le nombre d'observations pour chacun des niveaux des deux variables indépendantes (sexe et niveau de déficience intellectuelle) apparaît aussi dans ce tableau.

Tableau 16.5.
Plan factoriel d'analyse

	Valeur des catégories		N
Sexe	1	Féminin	66
	2	Masculin	76
Niveau de déficience intellectuelle	1	Légère	33
	2	Moyenne	55
	3	Grave/Profonde	54

Le tableau 16.6 présente les statistiques descriptives (moyennes, écarts types) pour chacune des variables dépendantes à l'étude en fonction des variables indépendantes (sexe et niveau de déficience intellectuelle) ainsi que pour l'ensemble des jeunes.

Tableau 16.6.
Statistiques descriptives en fonction du sexe et du niveau de déficience intellectuelle (extrait du Tableau SPSS pour les deux premières variables dépendantes)

	Sexe	Niveau de déficience intellectuelle	Moyennes	Écart-type	N
Comportements perturbateurs/ antisociaux	Féminin	Légère	12,5000	9,41970	14
		Moyenne	13,4375	11,23341	32
		Grave/profonde	12,5500	7,66726	20
		Total	12,9697	9,76331	66
	Masculin	Légère	14,6316	11,55283	19
		Moyenne	10,7391	9,12647	23
		Grave/profonde	12,2353	7,98506	34
		Total	12,3816	9,30801	76
	Total	Légère	13,7273	10,59561	33
		Moyenne	12,3091	10,40098	55
		Grave/profonde	12,3519	7,79733	54
		Total	12,6549	9,49285	142
Comportements centrés sur soi	Féminin	Légère	7,5714	4,71845	14
		Moyenne	8,7813	8,87815	32
		Grave/profonde	15,3000	10,42315	20
		Total	10,5000	9,17564	66
	Masculin	Légère	12,8947	12,15591	19
		Moyenne	8,6522	7,19601	23
		Grave/profonde	15,0882	10,63262	34
		Total	12,5921	10,39253	76
	Total	Légère	10,6364	9,96500	33
		Moyenne	8,7273	8,14556	55
		Grave/profonde	15,1667	10,45701	54
		Total	11,6197	9,86701	142

Le tableau des résultats de la MANOVA est ensuite disponible (voir tableau 16.7). Comme le devis de recherche à l'étude est un plan factoriel 2 (Sexe) × 3 (Niveau de déficience intellectuelle), les résultats obtenus à l'effet d'interaction sont habituellement interprétés en premier. D'après les résultats du tableau 16.7, l'effet d'interaction Sexe*DI n'est pas significatif selon le lambda de Wilks, car la probabilité associée à celui-ci est supérieure à 0,05. Il en est de même pour les trois autres tests (trace de Pillai, trace d'Hotelling et plus grande valeur propre de Roy).

Étant donné que l'effet d'interaction n'est pas significatif, il s'agit maintenant de vérifier si les effets principaux liés au sexe et au niveau de déficience intellectuelle le sont. Selon le tableau 16.7, les résultats à ces deux effets sont significatifs et, selon les valeurs des η^2 partiels associés à la statistique lambda de Wilks, ils expliquent respectivement 10,9 et 9,8 % de la variance.

Tableau 16.7.
Résultats de la MANOVA

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	,741	75,512[b]	5,000	132,000	,000	,741
	Wilks' Lambda	,259	75,512[b]	5,000	132,000	,000	,741
	Hotelling's Trace	2,860	75,512[b]	5,000	132,000	,000	,741
	Roy's Largest Root	2,860	75,512[b]	5,000	132,000	,000	,741
Sexe	Pillai's Trace	,109	3,245[b]	5,000	132,000	,009	,109
	Wilks' Lambda	,891	3,245[b]	5,000	132,000	,009	,109
	Hotelling's Trace	,123	3,245[b]	5,000	132,000	,009	,109
	Roy's Largest Root	,123	3,245[b]	5,000	132,000	,009	,109
DI	Pillai's Trace	,191	2,811	10,000	266,000	,002	,096
	Wilks' Lambda	,813	2,877[b]	10,000	264,000	,002	,098
	Hotelling's Trace	,225	2,941	10,000	262,000	,002	,101
	Roy's Largest Root	,198	5,265[c]	5,000	133,000	,000	,165
Sexe * DI	Pillai's Trace	,083	1,146	10,000	266,000	,328	,041
	Wilks' Lambda	,918	1,151[b]	10,000	264,000	,325	,042
	Hotelling's Trace	,088	1,155	10,000	262,000	,322	,042
	Roy's Largest Root	,077	2,040[c]	5,000	133,000	,077	,071

a : Design Intercept + sexe + DI + sexe * DI.
b : Exact statistic.
c : The statistic is an upper bound on F that yields a lower bound on the significance level.

Conséquemment, nous pouvons conclure que globalement, il y a des différences dans les problèmes de comportement selon le sexe des jeunes ainsi que leur niveau de déficience intellectuelle. La question est maintenant de savoir si ces différences concernent un ou des types de comportement en particulier.

8.3. Calcul de multiples ANOVA

L'une des procédures pour établir quelles variables dépendantes contribuent à la différence entre les groupes consiste à réaliser une ANOVA pour chacune d'elles. D'après le tableau 16.8, on remarque que l'effet sexe se manifeste uniquement à l'égard de l'anxiété et qu'il explique près de 4 % de la variance de la variable dépendante, [$F_{(1,\ 136)}$ = 5,028, p = 0,027, η_p^2 = 0,036]. En consultant les moyennes (figure 16.4), on remarque que les filles présentent davantage de comportements anxieux que les garçons. Cependant, il faut noter que si la correction de Bonferroni était appliquée, cet effet n'aurait pas été jugé significatif (0,05/5 = 0,01).

Tableau 16.8.
Résultats aux ANOVA

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	Comportements perturbateurs/ antisociaux	184,793[a]	5	36,959	,401	,847	,015
	Comportements centrés sur soi	1 400,625[b]	5	280,125	3,091	,011	,102
	Troubles de la communication	88,282[c]	5	17,656	1,111	,358	,039
	Anxiété	120,527[d]	5	24,105	2,157	,062	,073
	Problèmes sociaux	114,827[e]	5	22,965	2,198	,058	,075
Intercept	Comportements perturbateurs/ antisociaux	20 813,147	1	20 813,147	226,062	,000	,624
	Comportements centrés sur soi	16 762,128	1	16 762,128	184,934	,000	,576
	Troubles de la communication	5 007,110	1	5 007,110	315,015	,000	,699
	Anxiété	2 531,684	1	2 531,684	226,539	,000	,625
	Problèmes sociaux	2 193,428	1	2 193,428	209,929	,000	,607
Sexe	Comportements perturbateurs/ antisociaux	2,793	1	2,793	,030	,862	,000
	Comportements centrés sur soi	89,234	1	89,234	,985	,323	,007
	Troubles de la communication	17,672	1	17,672	1.112	,294	,008
	Anxiété	56,195	1	56,195	5,028	,027	,036
	Problèmes sociaux	2,836	1	2,836	,271	,603	,002

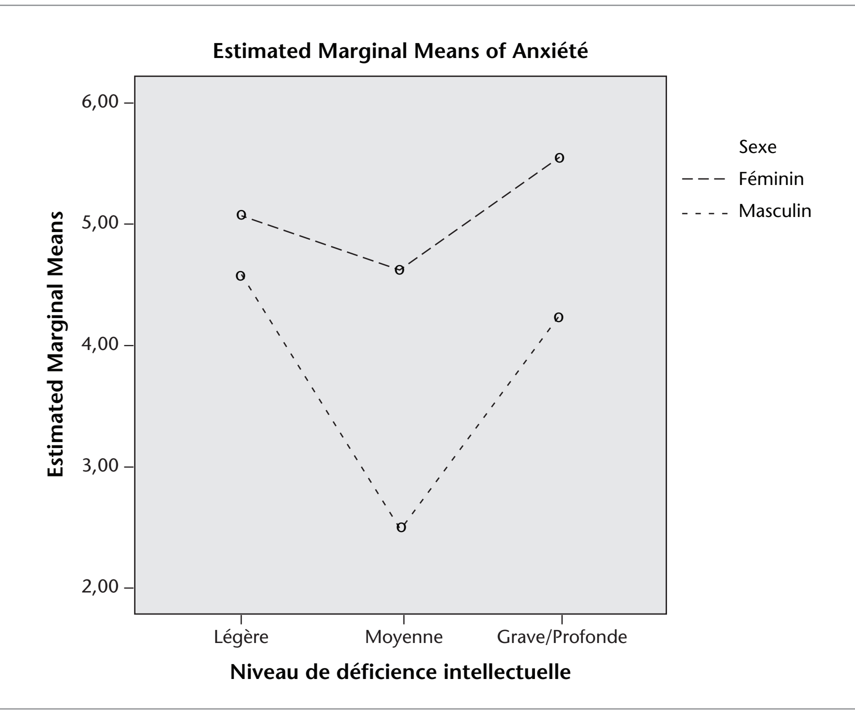
(*suite*)

Tableau 16.8. (*suite*)

DI	Comportements perturbateurs/ antisociaux	45,925	2	22,963	,249	,780	,004
	Comportements centrés sur soi	1 153,687	2	576,843	6,364	,002	,086
	Troubles de la communication	27,349	2	13,674	,860	,425	,012
	Anxiété	56,231	2	28,115	2,516	,085	,036
	Problèmes sociaux	48,969	2	24,484	2,343	,100	,033
Sexe * DI	Comportements perturbateurs/ antisociaux	119,912	2	59,956	,651	,523	,009
	Comportements centrés sur soi	185,621	2	92,810	1,024	,362	,015
	Troubles de la communication	58,723	2	29,362	1,847	,162	,026
	Anxiété	14,117	2	7,058	,632	,533	,009
	Problèmes sociaux	75,791	2	37,896	3,627	,029	,051
Erreur	Comportements perturbateurs/ antisociaux	12 521,298	136	92,068			
	Comportements centrés sur soi	12 326,839	136	90,639			
	Troubles de la communication	2 161,697	136	15,895			
	Anxiété	1 519,867	136	11,175			
	Problèmes sociaux	1 420,983	136	10,448			
Total	Comportements perturbateurs/ antisociaux	35 447,000	142				
	Comportements centrés sur soi	32 900,000	142				
	Troubles de la communication	7 891,000	142				
	Anxiété	4 330,000	142				
	Problèmes sociaux	3 913,000	142				
Total corrigé	Comportements perturbateurs/ antisociaux	12 706,092	141				
	Comportements centrés sur soi	13 727,465	141				
	Troubles de la communication	2 249,979	141				
	Anxiété	1 640,394	141				
	Problèmes sociaux	1 535,810	141				

a : R Squared =,015 (Adjusted R Squared = –,022)
b : R Squared =,102 (Adjusted R Squared = –,069)
c : R Squared =,039 (Adjusted R Squared = –,004)
d : R Squared =,073 (Adjusted R Squared = –,039)
e : R Squared =,075 (Adjusted R Squared = –,041)

Figure 16.4.
Moyennes des comportements anxieux selon le niveau de déficience intellectuelle et le sexe



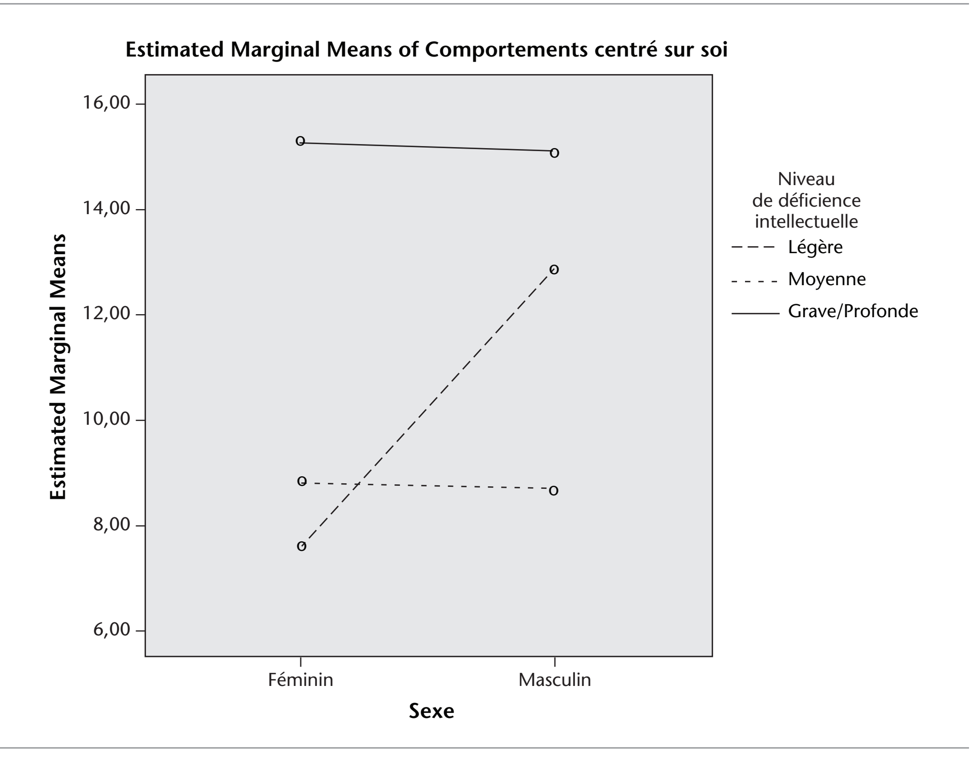
Si l'on considère maintenant les résultats des ANOVA concernant l'effet principal *Niveau de déficience intellectuelle* (voir le tableau 16.8), nous pouvons constater que des différences sont observées seulement en ce qui a trait aux comportements centrés sur soi, [$F_{(2,\ 136)} = 6{,}364$, $p = 0{,}002$, $\eta_p^2 = 0{,}086$]. De plus, cet effet explique environ 9 % de la variance de la variable dépendante. Comme cette variable comporte plus de deux catégories, nous devons recourir à des tests de comparaisons *a posteriori* pour situer les différences significatives entre les différents niveaux de déficience intellectuelle (tableau 16.9).

Tableau 16.9.
Comparaisons *a posteriori* pour les comportements centrés sur soi (extrait du Tableau SPSS)

			Mean Difference (I-J)	Std. Error	Sig.	95 % Confidence Interval	
						Lower Bound	Upper Bound
Comportements centrés sur soi							
Tukey HSD	Légère	Moyenne	1,9091	2,09633	,635	−3,0585	6,8767
		Grave/Profonde	−4,5303	2,10360	,083	−9,5151	,4545
	Moyenne	Légère	−1,9091	2,09633	,635	−6,8767	3,0585
		Grave/Profonde	−6,4394	1,82386	,002	−10,7613	−2,1175
	Grave/Profonde	Légère	4,5303	2,10360	,083	−.4545	9,5151
		Moyenne	6,4394	1,82386	,002	2,1175	10,7613
Scheffe	Légère	Moyenne	1,9091	2,09633	,661	−3,2792	7,0974
		Grave/Profonde	−4,5303	2,10360	,102	−9,7366	,6760
	Moyenne	Légère	−1,9091	2,09633	,661	−7,0974	3,2792
		Grave/Profonde	−6,4394	1,82386	,003	−10,9534	−1,9254
	Grave/Profonde	Légère	4,5303	2,10360	,102	−,6760	9,7366
		Moyenne	6,4394	1,82386	,003	1,9254	10,9534
Bonferroni	Légère	Moyenne	1,9091	2,09633	1,000	−3,1723	6,9905
		Grave/Profonde	−4,5303	2,10360	,099	−9,6293	,5687
	Moyenne	Légère	−1,9091	2,09633	1,000	−6,9905	3,1723
		Grave/Profonde	−6,4394	1,82386	,002	−10,8603	−2,0184
	Grave/Profonde	Légère	4,5303	2,10360	,099	−,5687	9,6293
		Moyenne	6,4394	1,82386	,002	2,0184	10,8603

Les différences statistiquement significatives sont celles où la probabilité associée au test (Sig.) est inférieure à 0,05. Indépendamment des tests réalisés (Tukey HSD, Scheffé ou Bonferroni), les résultats sont les mêmes. Les jeunes ayant un niveau de déficience intellectuelle grave/profond présentent davantage de comportements centrés sur eux que les jeunes ayant un niveau de déficience modéré (figure 16.5). Aucune autre différence n'est significative entre les niveaux de déficience intellectuelle, car les probabilités sont supérieures à 0,05.

Figure 16.5.
Moyennes des comportements centrés sur soi selon le sexe et le niveau de déficience intellectuelle



8.4. Présentation sommaire des résultats

Pour terminer cet exemple, voici comment les résultats de cette analyse pourraient être résumés. Bien entendu, il importe également de présenter une synthèse des résultats liés à la vérification des conditions d'application et, au besoin, de nuancer les résultats obtenus.

Une MANOVA 2 × 3 a été réalisée dans le but de vérifier les effets du sexe et du niveau de déficience intellectuelle des jeunes à l'égard de cinq types de problèmes de comportement : perturbateurs/antisociaux, centrés sur soi, troubles de la communication, anxiété et relations sociales. Les résultats de la MANOVA montrent des différences entre les filles et les garçons, [$F_{(5,\ 132)} = 3{,}245$, $p = 0{,}009$, Λ de Wilks = 0,891, $\eta_p^2 = 0{,}109$], ainsi qu'entre les divers niveaux de déficience intellectuelle [$F_{(10,\ 264)} = 2{,}877$, $p = 0{,}002$, Λ de Wilks = 0,813, $\eta_p^2 = 0{,}098$], mais l'interaction Sexe X Niveau

de déficience intellectuelle [$F_{(10, 264)} = 1{,}151$, $p = 0{,}325$, Λ de Wilks = 0,918] n'est pas significative. L'amplitude des deux effets principaux significatifs peut être qualifiée de moyenne selon l'interprétation de Cohen (1988), car elle se situe autour de 0,09. Afin de déterminer quels sont les types de comportements associés à ces effets significatifs, des ANOVA ont été réalisées sur chacun des types de problèmes de comportement.

Pour ce qui est du sexe, seuls les comportements anxieux s'avèrent significatifs [$F_{(1, 136)} = 5{,}028$, $p = 0{,}027$, $\eta_p^2 = 0{,}036$]. En effet, les filles manifestent davantage de comportements anxieux que les garçons. Or, cet effet, qui explique environ 4 % de la variance, ne serait pas jugé significatif si la correction de Bonferroni était appliquée.

Quant au niveau de déficience intellectuelle, les résultats aux ANOVA indiquent qu'ils se distinguent uniquement à l'égard des comportements centrés sur soi [$F_{(2, 136)} = 6{,}364$, $p = 0{,}002$, $\eta_p^2 = 0{,}086$] et que cet effet explique environ 9 % de ce type de comportements. Suivant les tests de comparaisons multiples (Tukey HSD, Scheffé ou Bonferroni), les enfants ayant un niveau de déficience intellectuelle grave/profond (M = 15,17) manifestent davantage ce type de comportements que les jeunes ayant un niveau de déficience modéré (M = 8,73). Les jeunes ayant un niveau de déficience légère (M = 10,64) ne se différencient pas significativement des deux autres groupes.

CONCLUSION

Ce chapitre poursuivait l'objectif de présenter, de façon accessible, un bref tour d'horizon de l'ANOVA et de la MANOVA, deux approches fréquemment utilisées aux fins de comparaison de moyennes. Bien qu'elles aient été développées initialement pour l'évaluation de l'effet de traitements dans le cadre d'expérimentations, elles sont souvent employées dans d'autres contextes de recherche.

Après avoir fait ressortir leur utilité ainsi que les différents plans d'analyse possible, la logique sous-jacente à ces deux méthodes d'analyse a été présentée. De plus, les principales conditions préalables à leur application dans le contexte de devis de recherche à groupes indépendants ont été abordées de même que certaines transformations de données ayant pour but de remédier à la violation de certaines de ces conditions. Toutefois, en pratique, il n'est pas toujours possible de respecter l'ensemble des postulats de base à ces analyses. Le travail d'exploration et de transformation de données requiert souvent beaucoup plus de temps que la réalisation de l'ANOVA ou de la MANOVA en tant que telle. Enfin, les grandes lignes de la

démarche d'analyse liées à l'application de l'ANOVA et de la MANOVA ont été illustrées au moyen d'un exemple. Cependant, la présentation de celui-ci a été limitée par certaines contraintes inhérentes aux résultats obtenus et aux procédures disponibles dans le logiciel SPSS.

En conclusion, le lecteur doit garder à l'esprit qu'il n'a pas été possible de couvrir dans cette présentation l'ensemble des possibilités d'analyse offertes par l'ANOVA et la MANOVA, car plusieurs chapitres y sont habituellement consacrés dans les ouvrages statistiques. Les plans à mesures répétées, les plans mixtes, l'analyse de covariance y ont été effleurés et leurs conditions d'application spécifiques (p. ex., sphéricité, homogénéité des plans de régression) n'ont pas été traitées. De plus, les limites imposées par les procédures disponibles dans les logiciels d'analyse statistique ont guidé, en partie, le choix des contenus présentés qui témoignent aussi des habitudes des usagers. Enfin, le lecteur intéressé par ces méthodes d'analyse est invité à consulter les écrits cités en références pour approfondir ses connaissances.

RÉFÉRENCES

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3e éd., New York, John Wiley & Sons.

BARNETT, V. et LEWIS, T. (1978). *Outliers in Statistical Data,* New York, John Wiley & Sons.

BLAIS, M. R., SABOURIN, S., BOUCHER, C. et VALLERAND, R. J. (1990). « Toward a motivational model of couple happiness », *Journal of Personality and Social Psychology*, vol. 59, no 5, p. 1021-103, <doi: 10.1037/0022-3514.59.5.1021>.

BOCK, R. D. (1975). *Multivariate Statistical Methods in Behavioural Research*, New York, McGraw-Hill.

BRAY, J. H. et MAXWELL, S. E. (1982). « Analyzing and interpreting significant MANOVA », *Review of Educational Reserach*, vol. 52, no 3, p. 340-367, <doi: 10.2307/1170422>.

COHEN, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2e éd., Hillsdale, Lawrence Erlbaum Associates.

COHEN, J. et COHEN, P. (1983). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences,* Hillsdale, Lawrence Erlbaum Associates.

COLE, D. A., MAXWELL, S. E., ARVEY, R. et SALAS, E. (1994). « How the power of MANOVA can both increase and decrease as a function of the intercorrelations among the dependent variables », *Psychological Bulletin*, vol. 115, no 3, p. 465-474, <doi: 10.1037/0033-2909.115.3.465>.

COOK, R. et WEISBERG, S. (1982). *Residuals and Influence in Regression*, New York, Chapman and Hall.

Einfeld, S. L. et Tonge, J. T. (2002). *Manual for the Developmental Behaviour Checklist second edition : Primary carer version (DBC-P) & Teacher version (DBC-T)*, Floride, Psychological Assessment Resources Inc.

Enders, C. K. (2010). *Applied Missing Data Analysis*, New York, Guilford Press.

Faraway, J. J. (2005). *Linear Models with R*, Boca Raton, Chapman & Hall/CRC.

Fisher, R. A. (1925). *Statistical Methods for Research Workers*, New York, Hafner.

Fisher, R. A. (1935). *The Design of Experiments*, New York, Hafner.

Fisher, R. A. (1956). *Statistical Methods and Scientific Inference*, New York, Hafner.

Fox, J. (2008). *Applied Regression Analysis and Generalized Linear Models*, 2e éd., Los Angeles, Sage.

Glass, G. V. et Hopkins, K. (1984). *Statistical Methods in Education and Psychology*, Englewood Cliffs, Prentice-Hall.

Glass, G. V., Peckham, P. D. et Sanders, J. R. (1972). « Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance », *Review of Educational Research*, vol. 42, nº 3, p. 237-288, <doi : 10.2307/1169991>.

Glèlè Kakaï R., Sodjinou E. et Fonton H. N. (2006). *Conditions d'application, des méthodes statistiques paramétriques : applications sur ordinateur,* Cotonou, Bénin, Bibliothèque Nationale.

Hair, J. F., Black, W. C., Babin, B. J. et Anderson, R. E. (2010). *Multivariate Data Analysis*, 7e éd., Upper Saddle River, Prentice Hall.

Hakstian, A. R., Roed, J. C. et Lind, J. C. (1979). « Two-sample T–2 procedure and the assumption of homogeneous covariance matrices », *Psychological Bulletin*, vol. 86, nº 6, p. 1255-1263, <doi : 10.1037/0033-2909.86.6.1255>.

Harris, R. J. (1975). *A Primer of Multivariate Statistics*, New York, Academic Press.

Hartley, H. O. (1950). « The use of range in analysis of variance », *Biometrika,* vol. 37, nºs 3-4, p. 271-280, <doi : 10.2307/2332380>.

Hotelling, H. (1931). « The generalization of Student's ratio », *Annals of Mathematical Statistics*, vol. 2, p. 360-378.

Hotelling, H. (1936). « Relation between two sets of variates », *Biometrika*, vol. 28, p. 321-377.

Howell, D. C. (2008). *Méthodes statistiques en sciences humaines*, 2e éd., Bruxelles, De Boeck.

Huitema, B. E. (1980). *The Analysis of Covariance and Alternatives*, New York, John Wiley & Sons.

Kendall, M. G. et Stuart, A. (1963). *The Advanced Theory of Statistics, Volume 1 : Distribution Theory*, 2e éd., New York, Hafner.

Kenny, D. A. et Judd, C. M. (1986). « Consequences of violating the independence assumption in analysis of variance », *Psychological Bulletin*, vol. 99, nº 3, p. 422-431, <doi : 10.1037/0033-2909.99.3.422>.

Kruskal, W. H. (1960). « Some remarks on wild observations », *Technometrics*, vol. 2, nº 1, p. 1-3, <doi : 10.2307/1266526>.

LACHANCE, L., RICHER, L., CÔTÉ, A. et TREMBLAY, K. N. (2010). *Étude des facteurs associés à la détresse psychologique des mères et des pères d'un enfant ayant une déficience intellectuelle et de leurs besoins en matière de services selon les caractéristiques de l'enfant*, Chicoutimi, Université du Québec à Chicoutimi.

LEHMANN, E. L. (2011). *Fisher, Neyman, and the Creation of Classical Statistics*, New York, Springer.

LILLIEFORS, H. W. (1967). « On the Kolmogorov-Smirnov Test for normality with mean and variance Unknown », *Journal of the American Statistical Association*, vol. 62, n° 318, p. 399-402, <doi : 10.2307/2283970>.

MCCULLAGH, P. et NELDER, J. A. (1989). *Generalized Linear Models*, 2e éd., Londres, Chapman & Hall.

NETER, J., WASSERMAN, W. et KUTNER, M. H. (1985). *Applied Linear Statistical Models. Regression, Analysis of Variance and Experimental Designs*, 2e éd., Homewood, Irwin.

NISHISATO, S. (2004). Dual scaling. Dans D. Kaplan (dir.), *The SAGE Handbook of Quantitative Methodology for the Social Sciences,* Thousand Oaks, Sage, p. 3-24.

OLSON, C. L. (1976). « On choosing a test statistic in multivariate analysis of variance », *Psychological Bulletin*, vol. 83, p. 579-586, <doi : 10.1037/0033-2909.83.4.579>.

RAUDENBUSH, S. W. et BRYK, A. S. (2002). *Hierarchical Linear Models. Applications and Data Analysis Methods*, 2e éd., Thousand Oaks, Sage.

ROBERT, C. P. (2007). *The Bayesian choice. From Decision-theoretic Foundations to Computational Implementation*, 2e éd., New York, Springer.

SAS INSTITUTE. (1990). *SAS/STAT User's guide. Volume 1*, 4e éd., Carey, SAS Institute.

SEARLE, S. R. (1971). *Linear Models*, New York, John Wiley & Sons.

SNIJDERS, T. et BOSKER, R. (1999). *Multilevel Analysis : An Introduction to Basic and Advanced Multilevel Modeling*, Thousand Oaks, Sage.

STEVENS, J. (1986). *Applied Multivariate Statistics for the Social Sciences*, Hillsdale, Lawrence Erlbaum Associates Inc.

STEVENS, J. P. (2009). *Applied Multivariate Statistics for the Social Sciences,* 5e éd., New York, Routledge/Taylor & Francis Group.

TABACHNICK, B. G. et FIDELL, L. S. (2013). *Using Multivariate Statistics*, 6e éd., Boston, Pearson Education.

THISSEN, D. et WAINER, H. (2001). *Test Scoring*, Mahwah, Lawrence Erlbaum Associates.

TOOTHAKER, L. E. (1993). *Multiple Comparison Procedures*, Newbury Park, Sage.

VON EYE, A. et BOGAT, G. A. (2004). « Testing the assumption of multivariate normality », *Psychology Science*, vol. 46, n° 2, p. 243-258.

WILKS, S. S. (1932). « Certain generalizations in the analysis of variance », *Biometrika*, vol. 24, n° 3-4, p. 471-494.

WISHART, J. (1928). « The generalized product moment distribution in samples from a normal multivariate population », *Biometrika*, vol. 20A, p. 32-52.

ZWICK, R. (1985). « Nonparametric one-way multivariate analysis of variance : A computational approach based on the Pillai–Bartlett trace », *Psychological Bulletin*, vol. 97, n° 1, p. 148-152, <doi : 10.1037/0033-2909.97.1.148>.

CHAPITRE 17

L'ANALYSE DISCRIMINANTE
Une analyse permettant d'explorer les prédicteurs de la satisfaction au travail

Marjolaine Beaudry
Julie Ménard

FORCES

- Elle permet la découverte de dimensions sous-jacentes à la formation de groupes, en fonction de variables indépendantes mesurées.
- Elle peut être réalisée à la suite d'une MANOVA afin d'interpréter les résultats de façon plus approfondie.
- Elle permet l'analyse de groupes de tailles inégales.

LIMITES

- Elle ne permet pas d'inférer une relation de cause à effet entre les variables indépendantes et les groupes.
- Elle doit répondre à un nombre plus important de postulats de base que d'autres analyses semblables, telles que la régression logistique.
- Elle est très sensible aux données extrêmes.

Ce chapitre a pour objectif de se familiariser avec l'analyse discriminante, souvent méconnue et peu utilisée malgré son potentiel intéressant en analyses exploratoires. En effet, cette analyse peut être utile pour les chercheurs intéressés à comprendre les caractéristiques de différents groupes d'individus, car elle permet d'analyser les variables expliquant la formation de groupes, ainsi que de découvrir les dimensions sous-jacentes expliquant la formation de groupes. Ce chapitre présentera, en premier lieu, un bref historique de l'analyse discriminante. L'analyse discriminante sera ensuite définie, puis le type de questions de recherche auxquelles elle permet de répondre sera décrit. En outre, les postulats et notions importantes de l'analyse discriminante seront présentés. Enfin, un devis d'étude (Ménard et Brunet, 2011) dont les résultats ont été analysés à l'aide de l'analyse discriminante sera présenté afin de permettre au lecteur d'avoir un exemple concret de ce type d'analyse et ainsi se familiariser avec son interprétation.

1. ANALYSE DISCRIMINANTE

1.1. Bref historique et définition

L'analyse discriminante a été élaborée par R. Fisher dans les années 1930. Ce scientifique a développé cette technique statistique dans le but de mieux comprendre ce qui différenciait différents types d'iris (McLachlan, 2004). Il a ainsi mesuré plusieurs caractéristiques physiques de cette fleur afin de produire une combinaison linéaire pour les distinguer. Ces combinaisons linéaires sont aujourd'hui la base de l'analyse discriminante. Dans les années 1960, l'analyse discriminante a été utilisée dans le domaine des finances afin de repérer les entreprises à risque de faillite (Altman, 1968). Peu à peu, le domaine du marketing a aussi utilisé cette analyse afin de créer des profils de consommateurs. Aujourd'hui, l'analyse discriminante est utilisée dans plusieurs domaines, tels que les sciences sociales et la santé, afin de comprendre les caractéristiques de personnes appartenant à différents groupes (Huberty et Olejnik, 2006).

De façon plus précise, l'analyse discriminante est une technique statistique multivariée qui permet de déterminer quelles sont les variables associées à la formation de groupes, et d'observer lesquelles de ces variables permettent le mieux d'assigner un sujet à un groupe ou à un autre. Ainsi, dans l'analyse discriminante, les variables indépendantes (aussi nommées prédicteurs) sont les mesures prises chez chacun des sujets, et la variable

dépendante correspond aux groupes à l'étude. La variable dépendante pourra être dichotomique s'il n'y a que deux groupes, ou catégorielle s'il y a plusieurs groupes.

Prenons l'exemple d'un échantillon de travailleurs qui ont vécu une invalidité au travail. Parmi ces travailleurs, il est possible de différencier trois groupes : un premier groupe n'ayant pas réintégré le travail, un deuxième ayant réintégré le travail chez le même employeur, et un troisième ayant réintégré le marché du travail, mais chez un employeur différent. Un chercheur pourrait être intéressé à savoir ce qui différencie ces groupes de travailleurs à partir d'un ensemble de variables indépendantes (p. ex., les relations qu'ils entretiennent avec leurs collègues, leur anxiété face au retour au travail, leur niveau socioéconomique, etc.), dans un éventuel but de déterminer les facteurs qui sont liés à la réintégration au travail. L'analyse discriminante permettra non seulement de comprendre plus en profondeur ce qui caractérise ces groupes, mais aussi de faire ressortir lesquelles de ces variables indépendantes permettent le mieux d'assigner un travailleur à un des trois groupes.

Dans la prochaine section, une comparaison entre l'analyse discriminante exploratoire et l'analyse discriminante confirmatoire sera présentée. Lorsque l'on ne connaît pas encore très bien quelles sont les caractéristiques propres à différents groupes, l'analyse discriminante exploratoire est plus appropriée. Dans ce cas, elle peut être comparée à l'analyse factorielle exploratoire (voir le chapitre 22 de cet ouvrage). Par contre, lorsque les variables associées à différents groupes sont connues et que l'on tente de prédire l'appartenance d'un sujet à un groupe, l'analyse discriminante confirmatoire est plus adéquate. Finalement, à la fin de la section, la fonction de classification de l'analyse discriminante sera présentée.

1.2. Types de questions de recherche auxquelles l'analyse discriminante permet de répondre

1.2.1. Analyse discriminante exploratoire

La question principale à laquelle l'analyse discriminante permet de répondre est : « Qu'est-ce qui explique l'appartenance d'un sujet à un groupe ? » De façon plus exploratoire, l'analyse discriminante permet avant tout de faire ressortir les liens entre des variables indépendantes mesurées dans le cadre d'une étude et la variable dépendante (les groupes) (Huberty et Olejnik, 2006). Cette étape, dite plus exploratoire de l'analyse discriminante, peut être comparée à l'analyse factorielle exploratoire. En effet, l'analyse factorielle

exploratoire permet notamment de faire ressortir des facteurs à partir d'items d'un questionnaire. Dans le cas de l'analyse discriminante, celle-ci permet de faire ressortir les dimensions sous-jacentes propres aux groupes à partir de l'analyse de variables indépendantes.

Afin de faire ressortir ces dimensions, on pourrait alors se poser les questions suivantes : « Est-ce que les individus de ces groupes partagent des caractéristiques communes ? » et « En quoi ces groupes sont-ils similaires ou différents ? ». Tout comme pour l'analyse factorielle, il est nécessaire que le chercheur se base sur l'état actuel des connaissances dans le domaine d'expertise à l'étude pour interpréter ces similitudes et différences propres à l'appartenance à chaque groupe (Tabachnick, Fidell et Osterlind, 2001). Il sera ainsi possible d'avoir une meilleure connaissance des caractéristiques qui peuvent expliquer la distinction entre les groupes. Par la suite, d'autres variables indépendantes en lien avec celles précédemment trouvées pourront alors être ajoutées et explorées afin d'expliquer la formation des groupes de façon plus précise.

1.2.2. Analyse discriminante confirmatoire

Lorsque les variables associées à chacun des groupes sont connues *a priori* dans la littérature scientifique ou encore lorsqu'une MANOVA (voir le chapitre 16 de cet ouvrage) a été effectuée afin de faire ressortir ces variables, l'analyse discriminante permet de déterminer lesquelles de ces variables contribuent à expliquer la séparation des groupes de façon optimale (Huberty et Olejnik, 2006). Le chercheur peut alors effectuer une analyse discriminante de façon confirmatoire avec plusieurs prédicteurs (ou variables indépendantes) et ensuite retirer de l'analyse ceux qui sont jugés comme étant non significatifs pour expliquer ce qui distingue les groupes. Ce type d'analyse est souvent exécuté à la suite d'une MANOVA. En effet, bien que la MANOVA permette de découvrir les variables associées à différents groupes ainsi que l'existence de relations entre ces variables, elle ne permet pas en soi de déterminer quelles variables parmi celles incluses dans l'analyse de variance expliquent le plus significativement la discrimination entre les groupes (Field, 2009). L'analyse discriminante devient alors intéressante, car elle permet d'explorer une à une les relations entre les variables et ainsi de voir si leur combinaison permet de prédire l'appartenance aux groupes. En somme, la MANOVA permet d'observer quelles variables sont associées aux différents groupes, tandis que l'analyse discriminante permet d'approfondir l'analyse en testant les combinaisons des variables dans le but de faire ressortir les meilleures combinaisons liées à la variable dépendante (le ou les groupes).

1.2.3. Classification

L'analyse discriminante permet aussi de faire une classification des sujets en groupes distincts. La procédure de classification de l'analyse discriminante permet d'observer si les sujets sont classés adéquatement dans leur groupe respectif à partir des variables indépendantes qui ont été mesurées. Cette classification permet aussi de déceler s'il existe des erreurs de classification parmi les participants à l'étude. Par exemple, se pourrait-il qu'il y ait une mauvaise classification de travailleurs en invalidité qui reçoivent un diagnostic de dépression ? Est-il possible que certains soient assignés au groupe « dépression » plutôt qu'au groupe « épuisement professionnel » et vice versa ? L'analyse discriminante permet ainsi d'observer si les variables indépendantes utilisées pour classer les travailleurs en invalidité ont été en mesure d'assigner adéquatement les travailleurs dans le bon groupe.

1.3. Notions de l'analyse discriminante

1.3.1. Fonction discriminante

Afin de faire ressortir les différences entre les groupes, l'analyse discriminante produit des combinaisons linéaires des variables indépendantes qui permettent de séparer les groupes de façon optimale, et ce, en prenant en considération les liens entre ces variables. Cette combinaison linéaire, appelée *fonction discriminante*, permettra de faire ressortir les dimensions sous-jacentes qui composent les différents groupes. Les fonctions discriminantes sont calculées à partir d'équations discriminantes, dont la forme est la suivante :

$$D = d_1 X_1 + d_2 X_2 + \cdots + a \quad (1)$$

où

- D = fonction discriminante ;
- d = coefficients discriminants non standardisés (comparables au coefficient bêta de la régression) ;
- X = score du sujet sur cette variable indépendante ;
- a = constante.

Cette équation permet de séparer les groupes en combinant les variables indépendantes de façon linéaire et en assignant un poids à chaque variable en fonction de ses coefficients discriminants. Ces coefficients correspondent aux *coefficients de discrimination canonique non standardisés* dont il sera question plus loin. L'équation discriminante doit être en

mesure d'établir les différences entre les groupes de façon optimale. Ainsi, plus le résultat d'une fonction discriminante est élevé, plus elle permettra de discriminer les groupes de façon optimale (Burns et Burns, 2008). En ce sens, cette fonction discriminante peut être comparée à une régression linéaire représentant une combinaison de variables indépendantes permettant de prédire une variable dépendante. La différence entre la régression (voir le chapitre 18 de cet ouvrage) et la fonction discriminante est que cette dernière permet de *prédire un groupe d'appartenance* tandis que la régression linéaire permet de *prédire une variable dépendante ayant un score continu*.

En outre, le nombre de fonctions discriminantes définies permet de savoir sous combien d'angles différents il est possible d'interpréter l'appartenance des sujets aux groupes (McLachlan, 2004). Ce nombre correspond au plus petit résultat entre le nombre de variables indépendantes, ou 1 – le nombre de groupes. Par exemple, dans une étude comportant trois variables indépendantes et deux groupes, une seule fonction discriminante sera nécessaire. Il est relativement simple de discriminer deux groupes. Par contre, l'interprétation devient plus complexe lorsqu'il y a présence de plus de deux groupes (Tabachnick *et al.*, 2001). En effet, il est possible d'observer une seule combinaison de variables pour expliquer une différence entre deux groupes. Par contre, il peut exister plus d'une façon d'expliquer la différence entre trois groupes.

Le nuage de points tel que représenté à la figure 17.1 permet de visualiser rapidement les fonctions discriminantes possibles ainsi que leur capacité à bien discriminer les groupes. Les fonctions discriminantes sont alors représentées par les axes, et les groupes, par leurs barycentres. Ces barycentres correspondent aux moyennes des groupes sur la fonction discriminante.

L'exemple de la figure 17.1 représente les trois groupes de travailleurs ayant vécu une invalidité au travail. Tous ont été analysés à partir des mêmes variables indépendantes, mais la combinaison de ces variables a permis de faire ressortir deux fonctions discriminantes. Dans cet exemple, la fonction discriminante 1 (lire la figure de façon verticale) permet de différencier les travailleurs ayant réintégré le travail chez le même employeur des travailleurs ayant réintégré le travail chez un autre employeur ou n'ayant pas réintégré le travail. Pour ce qui est de la fonction discriminante 2 (lire la figure de façon horizontale), celle-ci permet de discriminer le groupe n'ayant pas réintégré le marché du travail des deux autres groupes, ceux ayant réintégré le marché du travail (chez le même employeur ou non). Il s'agira, à ce moment, d'établir ce qui est commun et ce qui différencie ces deux combinaisons afin de bien interpréter la séparation des groupes.

Cette représentation graphique permet donc d'observer rapidement si les fonctions discriminantes permettent de bien discriminer les groupes et comment ils se distinguent entre eux.

Figure 17.1.
Barycentres de trois groupes de travailleurs sur les fonctions discriminantes obtenues à la suite d'une analyse discriminante



1.3.2. Équation de classification

En ce qui concerne la classification, celle-ci est réalisée à l'aide d'équations de classification créées pour chaque groupe. L'équation de classification prend la forme suivante :

$$C = c_1X_1 + c_2X_2 + \cdots + a \qquad (2)$$

De la même façon que pour l'équation de la fonction discriminante, cette équation fournit un score de classification (C) pour chaque groupe à partir d'un coefficient de classification (*c*) pour chaque variable indépendante (X), ainsi qu'une constante (*a*).

Une équation de classification est donc créée pour chacun des groupes. Lorsqu'un nouveau participant est ajouté à l'échantillon, un score de classification lui est assigné pour chacun des groupes. Le participant est alors classé dans un groupe en fonction de son score obtenu dans l'équation de classification, soit le groupe ayant un score le plus semblable au sien. Ces équations seront aussi utiles afin d'observer s'il y a des erreurs de classification de sujets dans un échantillon donné. L'analyse discriminante offre aussi une interprétation de la classification des sujets grâce aux tableaux de sortie obtenus à la suite de l'analyse. Cette interprétation est fournie dans la section suivante.

1.4. Procédure pour l'analyse discriminante et l'interprétation des résultats

Avant de procéder à l'analyse discriminante, les postulats de base suivants doivent être vérifiés (Tabachnick *et al.*, 2001; Burns et Burns, 2008):

- les variables indépendantes doivent être distribuées normalement;
- les scores des variables indépendantes doivent être, autant que possible, indépendants entre eux. Le non-respect de ce postulat est parfois considéré moins sévèrement, toutefois il peut avoir un effet sur la puissance statistique de l'analyse;
- l'échantillon doit être sélectionné aléatoirement;
- les groupes doivent être formés naturellement, c'est-à-dire qu'il ne doit pas y avoir de manipulation de la part du chercheur pour constituer les groupes. De plus, ces groupes doivent être connus avant la collecte des données;
- les groupes doivent être exclusifs tout en permettant d'inclure tous les participants;
- il doit y avoir une homogénéité variance-covariance pour chaque fonction discriminante. En effet, lorsque deux fonctions discriminantes ou plus sont nécessaires à l'analyse, celles-ci doivent posséder une homogénéité de variance-covariance;
- la proportion de la taille des groupes doit être comparable à ce que l'on retrouve dans la population. Ainsi, à la meilleure connaissance du chercheur, il ne doit pas y avoir une surreprésentation d'un groupe lorsque cela ne représente pas la réalité de la distribution de ces groupes dans la population.

Dans le cas où la taille des groupes n'est pas représentative, ou lorsque les postulats de base ne peuvent pas être respectés, la régression logistique peut être une alternative à l'analyse discriminante (Field, 2009). En effet, cette analyse permet de prédire une variable dépendante dichotomique à partir d'un ensemble de variables indépendantes. Toutefois, les postulats de base sont plus souples pour la régression logistique que pour l'analyse discriminante.

Une fois les postulats de base validés, on pourra procéder à l'analyse discriminante en suivant les six étapes décrites ci-dessous.

1.4.1. Déterminer si les groupes peuvent être discriminés par les variables indépendantes mesurées

Cette étape permet de constater s'il existe un effet principal entre les variables indépendantes pour chaque groupe, c'est-à-dire si on observe des différences de moyennes des variables indépendantes entre les groupes à l'étude. S'il n'existe pas de différence de moyennes pour certaines variables d'un groupe à l'autre, celles-ci ne peuvent donc discriminer les groupes. Par exemple, dans le cas présenté ci-dessus portant sur la réintégration au travail, s'il n'y a pas de différence entre les groupes sur leur niveau socio-économique et sur leur anxiété face à leur retour au travail, on doit conclure que ces groupes sont similaires en fonction de ces variables. Toutefois, si les moyennes sont significativement différentes, on passe à l'étape suivante.

1.4.2. Observer les relations entre les variables indépendantes et les groupes

Cette étape permet d'approfondir la compréhension du rôle des variables indépendantes dans la discrimination des groupes. Pour ce faire, une analyse du tableau de matrice de covariance permet d'observer la variance partagée entre les variables pour chacun des groupes. De plus, ce tableau montre si les variables indépendantes sont reliées positivement ou négativement aux groupes. Cette étape n'indique toutefois pas le poids de chaque variable sur les différents groupes, simplement si les variables sont associées aux groupes. Il est ainsi possible de bien comprendre la composition des groupes et par la suite de mieux interpréter les résultats obtenus par l'analyse discriminante. Pour ce qui est de l'homogénéité de variance des groupes, celle-ci peut être démontrée par le test M de Box, présenté à la section 2.3.2.

1.4.3. Déterminer à quel point les fonctions discriminantes permettent d'expliquer la différence entre les groupes

Cette étape permet d'analyser les fonctions discriminantes créées par l'analyse discriminante. En premier lieu, on observe la valeur propre (*eigen value*) associée à chacune des fonctions discriminantes. Cette valeur est aussi associée à un pourcentage de variance expliquée pour chaque fonction discriminante sur le modèle total, c'est-à-dire la contribution de chaque fonction discriminante pour expliquer la séparation des groupes. En second lieu, toujours dans le tableau des valeurs propres, le coefficient de la corrélation canonique associée à chaque fonction discriminante est aussi considéré. Cette valeur, lorsqu'elle est mise au carré, représente le pourcentage de variance expliquée pour chaque fonction discriminante. Elle est comparable au R^2 de la régression.

1.4.4. Déterminer si les fonctions discriminantes sont significatives

Une fonction discriminante est dite significative lorsqu'elle permet de prédire l'appartenance à un groupe de façon plus précise que par le simple hasard. La significativité du coefficient du lambda de Wilks indique si les fonctions discriminantes sont significatives ou non. Lorsqu'il existe plus d'une fonction discriminante, le premier indice de significativité fourni correspond au modèle complet (c.-à-d. pour l'ensemble des fonctions discriminantes), et le deuxième, au niveau de significativité de la fonction suivante en retirant la contribution de la première. La première fonction discriminante trouvée est celle qui permet le mieux de discriminer les individus en groupes. La deuxième fonction discriminante permet alors d'expliquer les différences entre les groupes qui n'ont pas été expliquées par la première fonction.

1.4.5. Déterminer quelles variables indépendantes sont associées aux fonctions discriminantes, et de quelle manière

Pour cette étape, il s'agit tout d'abord d'observer les *coefficients canoniques standardisés* des variables indépendantes pour chacune des fonctions discriminantes. Ces coefficients renseignent sur la contribution de chaque variable à chacune des fonctions discriminantes, ainsi que sur sa capacité à prédire l'appartenance à un groupe. Le signe associé à chaque variable nous permet aussi de savoir si cette variable est associée positivement ou négativement à la fonction discriminante.

Ensuite, la matrice de structure (*structure matrix*) peut également renseigner sur la contribution de chaque variable indépendante aux fonctions discriminantes. Cette matrice indique la corrélation entre la variable et la fonction discriminante. Les corrélations de la matrice de

structure correspondent à des *indices de saturation discriminants* pouvant être comparés aux coefficients de saturation de l'analyse factorielle. Par ailleurs, lorsque plusieurs variables indépendantes sont corrélées à une seule fonction discriminante et que d'autres variables sont associées à une autre fonction discriminante, le chercheur tente de comprendre quelles caractéristiques sont communes au premier ensemble de variables et ce qui les différencie du deuxième groupe de variables. La matrice de structure permet ainsi d'interpréter chaque fonction discriminante de façon plus complète selon les variables qui y sont associées et selon celles qui ne le sont pas. Cette interprétation ne peut être conduite que par un chercheur ayant une bonne connaissance théorique des variables indépendantes et des groupes qui sont à l'étude.

Un troisième indice nous permet de compléter l'interprétation des fonctions discriminantes. Il s'agit des *coefficients canoniques discriminants*. Ces coefficients sont la version non standardisée des coefficients canoniques standardisés présentés plus haut. Ces versions non standardisées sont utilisées dans l'équation (1) permettant de calculer les fonctions discriminantes. Ils indiquent la contribution unique de chaque variable à chaque fonction discriminante, tout en contrôlant la contribution des autres variables.

Finalement, une observation des moyennes des groupes, appelées barycentres, de chaque fonction discriminante permet de bien discerner la différence entre les groupes par rapport aux fonctions discriminantes. Ces barycentres sont disponibles en format tableau, mais aussi de façon graphique, comme le nuage de points présenté à la figure 17.1.

1.4.6. Classifier

Si le but de l'analyse consiste à interpréter la classification des sujets, le tableau de classification permet d'observer si les variables indépendantes mesurées ont permis de classer les sujets adéquatement. Ces tableaux présentent une comparaison des sujets selon leur groupe d'origine et leur groupe prédit. Lorsque les sujets sont parfaitement classés, le résultat sera le même pour ces deux catégories (groupe d'origine et groupe prédit). Cette comparaison est aussi accompagnée d'un pourcentage de classification pour chacun des groupes.

1.5. Logiciel et matériel nécessaires pour exécuter l'analyse

La majorité des logiciels d'analyse statistique, tels que SPSS et SAS, permettent de réaliser des analyses discriminantes et offrent sensiblement les mêmes tableaux de sortie permettant une interprétation complète des résultats de l'analyse.

Afin de mieux comprendre comment l'analyse discriminante peut être utilisée, une étude dont les résultats ont été analysés avec l'analyse discriminante est présentée dans la section suivante.

2. APPLICATION CONCRÈTE DE L'ANALYSE DISCRIMINANTE POUR EXPLORER LES PRÉDICTEURS DE LA SATISFACTION AU TRAVAIL

Une étude transversale réalisée auprès de 360 gestionnaires du secteur public et parapublic a été effectuée dans le but de mieux comprendre ce qui distingue les gestionnaires satisfaits au travail des gestionnaires insatisfaits au travail. La recension des écrits effectuée révèle que la majorité des études sur la satisfaction au travail se sont intéressées aux facteurs organisationnels (Viswesvaran, Sanchez et Fisher, 1999), négligeant ainsi les prédicteurs individuels et interpersonnels. Dans cette étude, plusieurs prédicteurs ont été mesurés par l'intermédiaire de questionnaires autorapportés avec des échelles de type Likert : 1) la qualité des relations avec les collègues, mesurée par l'échelle de Ryff et Keyes (1995) ; 2) le sens du travail, mesuré par l'échelle de Steger *et al*. (2006) ; 3) l'authenticité cognitive (c.-à-d. une compréhension et une évaluation non biaisée du Soi) ; et 4) l'authenticité comportementale (c.-à-d. agir en concordance avec son vrai Soi et être vrai dans ses interactions et relations), ces deux derniers étant mesurés par l'échelle de Goldman et Kernis (2004). À partir de l'analyse de ces quatre prédicteurs, il sera possible d'observer s'il existe une ou plusieurs dimensions sous-jacentes qui expliquent la différence entre les gestionnaires qui rapportent être satisfaits et ceux qui rapportent être insatisfaits au travail. L'analyse discriminante nous permettra aussi d'observer si les prédicteurs utilisés permettent de bien classer les travailleurs, selon qu'ils sont satisfaits ou non.

2.1. Mode d'évaluation utilisé

Pour réaliser une analyse discriminante, plusieurs types de mesure peuvent être utilisés. Il peut s'agir de mesures autorapportées, comme dans l'étude citée en exemple, ou encore de mesures biographiques comme des caractéristiques du récit de vie, de mesures d'observation ou encore de mesures physiologiques. Ce qui importe, c'est que les prédicteurs soient de type continu et que la variable dépendante (le groupe d'appartenance) soit de nature catégorielle (deux catégories ou plus).

Dans le cadre de l'étude présentée, l'échelle de satisfaction de Diener (1984), une échelle d'accord de type Likert allant de 1 = fortement en désaccord à 7 = fortement en accord a été utilisée comme variable dépendante.

Elle a été transformée en échelle catégorielle aux fins de l'analyse. Ainsi, avant d'effectuer l'analyse discriminante, les gestionnaires qui ont répondu comme étant plutôt insatisfaits vis-à-vis de leur travail (1 à 3,5 sur 7) ont été classés dans le groupe des gestionnaires insatisfaits, alors que ceux qui ont répondu qu'ils étaient plutôt satisfaits (3,6 à 7 sur 7) ont été classés dans le groupe des gestionnaires satisfaits.

2.2. Présentation de la procédure d'analyse

En ce concerne le vocabulaire utilisé dans les logiciels d'analyse, la variable de groupe (*satisfaits* ou *non satisfaits*) correspond au critère de regroupement qualitatif numérique, et les prédicteurs s'apparentent aux *variables indépendantes*. Avant de pouvoir effectuer l'analyse, il est nécessaire de définir l'intervalle du critère de regroupement qualitatif numérique. Dans notre exemple, la valeur 0 a été attribuée aux gestionnaires qui appartiennent à la catégorie des *insatisfaits au travail*, et 1 à ceux qui appartiennent à la catégorie des *satisfaits au travail*.

Lorsque l'on procède à l'analyse en tant que telle, il est essentiel de débuter en déterminant les statistiques que l'on souhaite calculer. Généralement, la *moyenne*, le *test de Box*, les *coefficients de la fonction non standardisés* et la *matrice de corrélation intraclasse* sont nécessaires. Si le but de l'analyse est la classification, il est recommandé de se baser sur les probabilités *a priori* «*égales pour toutes les classes*», et d'utiliser la *matrice de covariance intraclasse*. Une fois l'analyse effectuée, il est suggéré d'enregistrer le résultat des analyses en fonction du *groupe d'affectation prévu* de même que les *scores discriminants*. Finalement, les tableaux de sortie permettent de procéder à l'interprétation des résultats, telle que décrite dans la section qui suit.

2.3. Interprétation des résultats

2.3.1. Discrimination des groupes à partir des prédicteurs

Tout d'abord, il est important de s'assurer que les postulats relatifs à la collecte de données ont été respectés, c'est-à-dire : 1) que les observations sont indépendantes entre elles (étude transversale) ; 2) que les données ont été échantillonnées aléatoirement (le recrutement a été fait par l'entremise de plusieurs associations de cadres, dont les membres se sont portés volontaires) ; 3) que les groupes ont été formés naturellement en fonction des réponses des participants et que l'appartenance à un groupe a été déterminée avant la collecte des données (le score attribué au groupe de travailleurs

satisfaits et non satisfaits a été déterminé *a priori*) ; 4) que les groupes sont exclusifs tout en permettant d'inclure tous les participants ; et 5) qu'il n'y a pas de surreprésentation de travailleurs non satisfaits au travail, ce qui ne correspondrait pas à la réalité québécoise (Fortin et Traoré, 2007). Une fois que ces postulats sont respectés, le premier critère à vérifier est d'observer s'il existe une différence significative entre les groupes concernant les prédicteurs de la fonction.

Le tableau 17.1 représente les tailles d'effectif de chaque groupe déterminé *a priori* ainsi que les moyennes et écarts-types de chaque prédicteur pour chacun des groupes. À première vue, on peut constater que la différence de moyennes entre les deux groupes pour le sens accordé au travail est la plus élevée (1,3857). Toutefois, il se peut que ce prédicteur ne soit pas le seul qui soit significativement différent pour chacun des groupes.

Tableau 17.1.
Statistiques de groupe

Satisfaction au travail		Moyenne	Écart-type	N valide (liste)	
				Non pondérées	Pondérées
NON	Relations positives avec les collègues	4,0437	0,72687	32	32,000
	Sens du travail	4,9625	0,99474	32	32,000
	Authenticité cognitive	3,0271	0,59014	32	32,000
	Comportements authentiques	3,5097	0,49412	32	32,000
OUI	Relations positives avec les collègues	4,7677	0,70135	325	325,000
	Sens du travail	6,3482	0,66446	325	325,000
	Authenticité cognitive	3,5921	0,55985	325	325,000
	Comportements authentiques	4,0299	0,46894	325	325,000
Total	Relations positives avec les collègues	4,7028	0,73251	357	357,000
	Sens du travail	6,2239	0,80318	357	357,000
	Authenticité cognitive	3,5414	0,58456	357	357,000
	Comportements authentiques	3,9833	0,49350	357	357,000

Afin de savoir si les différences de moyennes sont significatives pour chacun des prédicteurs, il est nécessaire d'observer les résultats du tableau 17.2. Celui-ci permet de s'assurer que les moyennes des groupes sont significativement différentes pour chacun des prédicteurs inclus dans l'analyse discriminante.

Tableau 17.2.
Test d'égalité des moyennes de groupe

	F	*ddl*1	*ddl*2	Signification
Relations positives avec les collègues	30,841	1	355	,000
Sens du travail	75,679	1	355	,000
Authenticité cognitive	29,385	1	355	,000
Authenticité comportementale	35,513	1	355	,000

Dans le tableau 17.2, on observe qu'il existe une différence significative de moyenne sur chacun des prédicteurs entre les gestionnaires qui sont satisfaits et ceux qui sont insatisfaits de leur travail. Ce tableau permet également de constater que le sens du travail semble être le prédicteur dont la moyenne diffère de façon la plus importante entre les groupes de par sa valeur de F (test de Fisher) plus élevée. En outre, les moyennes étant significativement différentes pour chaque prédicteur, il est alors possible de poursuivre l'interprétation de l'analyse discriminante.

2.3.2. Homogénéité variance-covariance

Le second critère à vérifier est le respect du postulat d'homogénéité de variance et covariance, essentiel pour l'analyse discriminante. Les déterminants logs et le M de Box présentés dans les tableaux 17.3 et 17.4 permettront de s'assurer que ce postulat est respecté.

Tableau 17.3.
Déterminants logs

Satisfaction au travail	Rang	Déterminant log
0,00	4	−4,196
1,00	4	−4,849
Intragroupes combinés	4	−4,744

Les rangs et logarithmes naturels des déterminants imprimés sont ceux des matrices de covariance du groupe.

Tableau 17.4.
M de Box

M de Box		16,978
F	Approximativement	1,617
	*ddl*1	10
	*ddl*2	12594,069
	Signification	0,095

Test de l'hypothèse nulle d'égalité de matrices de covariance des populations.

Pour s'assurer que le postulat d'homogénéité de variance est respecté, les déterminants logs doivent être comparables entre les groupes, et le test M de Box doit être non significatif. Le fait que ce test soit non significatif nous permet de conserver l'hypothèse nulle d'homogénéité de variance. Dans le cas présent, les déterminants logs sont de valeurs similaires (–4,196 et –4,849), et le test F du M de Box présente une valeur de $p = 0{,}095$. Les matrices de covariance ne sont donc pas différentes et les données de l'étude répondent au postulat d'homogénéité des variances. Les postulats de base sont ainsi respectés, et l'interprétation des résultats de l'analyse discriminante peut être effectuée.

2.3.3. Analyse des fonctions discriminantes

Comme mentionné précédemment, l'analyse discriminante crée une ou des fonctions discriminantes. Il est important de rappeler que cette fonction discriminante est une combinaison linéaire des prédicteurs à l'étude permettant de séparer les groupes de façon optimale. De plus, cette fonction discriminante nous permet de déterminer si la combinaison des prédicteurs permet de faire ressortir une dimension sous-jacente aux données de l'étude. Dans le cas présent, une seule fonction discriminante a été créée, car nous n'avons que deux groupes et le nombre de fonctions discriminantes correspond à $k - 1$, (k étant le nombre de groupe). Le tableau 17.5 nous renseigne sur cette fonction discriminante.

Tableau 17.5.
Valeurs propres

Fonction	Valeur propre	% de la variance	% cumulé	Corrélation canonique
1	0,359[a]	100,0	100,0	0,514

a. Les (1) premières fonctions discriminantes canoniques ont été utilisées pour l'analyse.

Comme il n'y a qu'une seule fonction discriminante, celle-ci explique 100 % de la variance du modèle. Dans un cas où plus d'une fonction discriminante émergerait de l'analyse, chacune d'entre elles pourrait expliquer un pourcentage de variance plus ou moins important, ce qui constitue un bon indicateur en soi afin de retenir la meilleure combinaison de prédicteurs.

Une corrélation canonique est aussi associée à la fonction discriminante (0,514 dans notre exemple). Cette corrélation peut être convertie en pourcentage de variance expliquée (aussi appelée taille de l'effet) lorsqu'elle est mise au carré. Ainsi, dans notre cas, nous pouvons affirmer que la fonction discriminante permet d'expliquer 26,4 % de la variation entre les groupes (c.-à-d. si les gestionnaires sont satisfaits ou insatisfaits au travail).

2.3.4. Degré de signification de la fonction discriminante

Par la suite, le tableau 17.6 nous indique si la fonction discriminante trouvée est significative ou non par l'intermédiaire du lambda de Wilks.

Tableau 17.6.
Lambda de Wilks

Test de la ou des fonctions	Lambda de Wilks	Chi-carré	*ddl*	Signification
1	0,736	108,281	4	0,000

Dans le cas présent, la fonction discriminante est significative ($p < 0{,}000$). Nous pouvons donc affirmer que la fonction discriminante, représentant la combinaison des prédicteurs, permet d'expliquer la séparation entre les groupes de façon significative.

2.3.5. Analyse du rôle des prédicteurs dans la fonction discriminante

Ces premières analyses ont permis de démontrer que les prédicteurs et la fonction discriminante permettent de bien distinguer les gestionnaires satisfaits au travail de ceux qui sont insatisfaits au travail. Il importe ensuite de comprendre plus en profondeur l'effet des prédicteurs sur la formation des groupes. Ces analyses permettent donc d'évaluer le poids de chacun de ces prédicteurs et de déterminer s'ils sont associés de façon positive ou négative à la formation des groupes. Afin d'y arriver, il est nécessaire d'observer les coefficients de chacun des prédicteurs. Le tableau 17.7 nous indique les coefficients des variables indépendantes sur la fonction discriminante canonique standardisée.

Tableau 17.7.
Coefficients de la fonction discriminante canonique standardisée

	Fonction 1
Relations positives avec ses collègues	0,263
Sens du travail	0,840
Authenticité cognitive	0,041
Authenticité comportementale	0,105

Chaque coefficient peut être interprété comme un coefficient bêta d'une régression. Ainsi, par l'observation de chaque coefficient, on observe la nature de la relation de la variable avec la fonction discriminante. On peut ainsi déterminer si une variable est associée positivement ou négativement à la fonction discriminante ainsi que son poids relatif. Dans le cas présent, nous pouvons observer que tous les prédicteurs de la fonction sont associés positivement à la fonction discriminante. On observe que le sens du travail contribue de façon beaucoup plus importante à la fonction discriminante (0,840), suivi des relations positives avec les collègues (0,263). Les deux types d'authenticité (0,105 pour l'authenticité comportementale et 0,041 pour l'authenticité cognitive) sont les prédicteurs qui contribuent le moins à la fonction discriminante. On comprend donc que pour distinguer les gestionnaires satisfaits des gestionnaires insatisfaits de leur travail, le sens que ceux-ci donnent à leur travail est très important. Ce résultat est congruent avec la théorie de Ryff et Keyes (2005), qui stipule que le sens du travail est fortement lié à la satisfaction au travail.

La matrice de structure présentée au tableau 17.8 nous indique, quant à elle, la corrélation entre les prédicteurs et la fonction discriminante. Ces corrélations, aussi appelées indices de saturation discriminants, peuvent être comparées à des indices de saturation utilisés pour une analyse factorielle exploratoire (voir le chapitre 22 de cet ouvrage). Ainsi, les prédicteurs ayant un indice discriminant plus élevé contribuent plus fortement à la séparation des groupes. Certains auteurs (Burns et Burns, 2008; Tabachnick *et al.*, 2001) considèrent que ces facteurs discriminants sont plus fiables que les coefficients standardisés de la fonction canonique discriminante pour attester de l'importance relative de chaque prédicteur sur la fonction discriminante. Ainsi, la variable de sens accordé au travail possède encore une fois la valeur la plus élevée (0,947). Pour ce qui est des trois autres prédicteurs, ceux-ci présentent des coefficients plus semblables les uns aux autres (0,528 pour l'authenticité comportementale, 0,492 pour les relations positives avec les autres, et 0,480 pour l'authenticité cognitive). Bien que ces indices discriminants soient plus faibles, ils doivent être considérés comme étant liés au

fait d'être satisfait ou non au travail. En effet, lorsqu'une variable présente un indice discriminant qui dépasse le seuil de 0,30 (même valeur seuil pour l'indice de saturation obtenue avec l'analyse factorielle, voir le chapitre 22 de cet ouvrage), celle-ci est considérée comme étant liée à la formation des groupes à l'étude. Ainsi, nous pouvons affirmer que le fait d'être satisfait ou non au travail pourrait être associé à deux dimensions sous-jacentes, la première étant associée au sens accordé au travail, et la deuxième regroupant les deux formes d'authenticité ainsi que les relations positives entretenues avec les collègues.

Tableau 17.8.
Matrice de structure

	Fonction 1
Sens du travail	0,947
Authenticité comportementale	0,528
Relations positives avec les autres	0,492
Authenticité cognitive	0,480

Les corrélations intragroupes combinées entre les variables discriminantes et les variables des fonctions discriminantes canoniques standardisées sont ordonnées par taille absolue des corrélations à l'intérieur de la fonction.

Finalement, le tableau 17.9 nous présente les coefficients de fonctions canoniques discriminantes non standardisés. Ces coefficients non standardisés sont aussi utilisés pour calculer la fonction discriminante (équation 1) présentée dans la première section de ce chapitre. Nous pouvons constater encore une fois que le sens accordé au travail est celui présentant le poids le plus important, suivi des relations positives avec les collègues.

Tableau 17.9.
Coefficients des fonctions discriminantes canoniques

	Fonction 1
Relations positives avec ses collègues	0,373
Sens du travail	1,200
Authenticité cognitive	0,074
Authenticité comportementale	0,224
(Constante)	–10,379

Coefficients non standardisés.

Les résultats de chacun des participants à l'équation discriminante établie à partir de ces coefficients non standardisés peuvent être sauvegardés dans la banque de données. Lorsque les options de sauvegarde ont été effectuées, les scores discriminants sont ainsi créés pour chaque participant. Pour aider à la visualisation, une moyenne (appelée barycentre) des scores pour chaque groupe peut être observée de façon graphique, comme proposé à la figure 17.1, ou sous forme de tableau (tableau 17.10).

Tableau 17.10.
Fonctions aux barycentres des groupes

Satisfaction au travail	Fonction 1
0,00	−1,774
1,00	0,175

Fonctions discriminantes canoniques non standardisées évaluées aux moyennes des groupes.

2.3.6. Classification des participants

Une deuxième utilité de l'analyse discriminante est la classification des participants au sein des différents groupes. Le tableau 17.11 nous renseigne sur l'efficacité de cette classification à l'aide de l'équation de classification.

Les lignes de ce tableau correspondent au véritable groupe d'appartenance de chaque répondant, tandis que les colonnes correspondent au groupe prédit par l'équation de classification. En comparant ces données, il est ainsi possible de déterminer le nombre de répondants qui ont été correctement classés et ceux qui ont été classés dans la mauvaise catégorie. Ainsi, pour les répondants insatisfaits (n = 35), 22 ont été classés dans la bonne catégorie, tandis que 10 ont été classés comme étant satisfaits au travail. Un pourcentage est aussi fourni par l'équation de classification, permettant ainsi une meilleure compréhension des données. Dans le même exemple, 68 % des répondants ont correctement été classés comme étant insatisfaits au travail, tandis que 31 % ont été incorrectement classés comme étant satisfaits au travail. On comprend donc que nos prédicteurs (variables indépendantes à l'étude) ont permis de bien classer la majorité des participants à notre étude, mais que d'autres prédicteurs sont nécessaires afin de permettre de prédire l'appartenance à chacune des catégories de gestionnaires, particulièrement pour les gestionnaires insatisfaits.

Tableau 17.11.
Résultats du classement

		Satisfaction au travail	Classe(s) d'affectation prévue(s)		Total
			0,00	1,00	
Original[a]	Effectif	0,00	22	10	32
		1,00	48	277	325
		Observations non classées	1	1	2
	%	0,00	68,8	31,3	100,0
		1,00	14,8	85,2	100,0
		Observations non classées	50,0	50,0	100,0
Validé-croisé[b, c]	Effectif	0,00	22	10	32
		1,00	49	276	325
	%	0,00	68,8	31,3	100,0
		1,00	15,1	84,9	100,0

[a] 85,2 % des observations originales classées correctement.
[b] La validation croisée n'est effectuée que pour les observations de l'analyse. Dans la validation croisée, chaque observation est classée par les fonctions dérivées de toutes les autres observations.
[c] 84,9 % des observations validées-croisées classées correctement.

La section validé-croisé permet aussi d'interpréter la classification, mais de façon plus fidèle. Elle nous informe sur l'adéquation de l'équation de classification quant à sa capacité à classer correctement les participants à l'étude, dans leur groupe d'appartenance prévu *a priori*. En outre, le validé-croisé indique si la combinaison des prédicteurs mesurés permettrait de classer correctement de nouveaux sujets dans une future étude. Cette procédure, aussi connue sous le nom de classification «*jackknifed*», consiste à dériver la classification d'un sujet à partir des données des autres sujets dans l'échantillon. Dans le cas présent, l'équation de classification a permis de classifier correctement 68,8% des gestionnaires insatisfaits et 84,9% des gestionnaires satisfaits au travail. On peut donc conclure que la fonction discriminante permet de bien classer nos gestionnaires, particulièrement les gestionnaires satisfaits de leur travail.

CONCLUSION

En somme, l'analyse discriminante est une procédure polyvalente permettant de tirer des conclusions variées sur l'appartenance de sujets à différents groupes. En effet, elle permet de déterminer quelles sont les variables permettant de discriminer les groupes de façon optimale, de découvrir les dimensions sous-jacentes à la formation de groupes et finalement d'effectuer une classification de sujets dans leur groupe d'appartenance. Le lecteur désirant approfondir certains aspects de la fonction discriminante peut se référer à l'ouvrage de McLachlan (2004), qui offre des explications détaillées des nombreuses fonctions de l'analyse discriminante.

RÉFÉRENCES

ALTMAN, E.I. (1968). « Financial ratios, discriminant analysis and the prediction of corporate bankruptcy », *The Journal of Finance*, vol. 23, nº 4, p. 589-609.

BURNS, R. et R.P. BURNS (2008). *Business Research Methods and Statistics Using SPSS*, Londres, Sage Publications.

DIENER, E. (1984). « Subjective well-being », *Psychological Bulletin*, vol. 95, p. 542-75.

DIENER, E., R.A. EMMONS, R.J. LARSEN et S. GRIFFIN (1985), « The satisfaction with life scale », *Journal of Personality Assessment*, vol. 49, nº 1, p. 71-75.

FIELD, A. (2009). *Discovering Statistics using SPSS*, Londres, Sage Publications.

FORTIN, É. et I. TRAORÉ (2007). *Les Québécois sont-ils satisfaits de leur vie et du cadre de vie dans lequel ils évoluent ?*, Québec, Zoom Santé, Santé et Bien-être, Institut de la statistique du Québec.

GOLDMAN, B.M. et M.H. KERNIS (2004). *The Authenticity Inventory*, Athens, University of Georgia.

HUBERTY, C J. et S. OLEJNIK (2006). *Applied MANOVA and Discriminant Analysis*, New York, John Wiley & Sons.

MCLACHLAN, G. (2004). *Discriminant Analysis and Statistical Pattern Recognition*, New York, John Wiley & Sons.

MÉNARD, J. et L. BRUNET (2011). « Authenticity and well-being in the workplace : A mediation model », *Journal of Managerial Psychology*, vol. 26, nº 4, p. 331-346.

RYAN, R.M. et E.L. DECI (2001). « On happiness and human potentials : A review of research on hedonic and eudaimonic well-being », *Annual Review of Psychology*, vol. 51, nº 1, p. 141-166.

RYFF, C.D. et C.L. KEYES (1995). « The structure of psychological well-being revisited », *Journal of Personality and Social Psychology*, vol. 69, nº 4, p. 719-727.

STEGER, M.F., P. FRAZIER, S. OISHI et M. KALER (2006). « The meaning in life questionnaire : Assessing the presence of and search for meaning in life », *Journal of Counseling Psychology*, vol. 53, p. 80-93.

TABACHNICK, B.G., L.S. FIDELL et S.J. OSTERLIND (2001). *Using Multivariate Statistics*, Boston, Allyn & Bacon/Pearson Education.

VISWESVARAN, C., J.I. SANCHEZ et J. FISHER (1999). « The role of social support in the process of work stress : A meta-analysis », *Journal of Vocational Behavior*, vol. 54, n° 2, p. 314-334.

CHAPITRE 18

LES RÉGRESSIONS LINÉAIRES ET LOGISTIQUES
Illustration de la prédiction de l'anxiété chez les enseignants

Nathalie Houlfort
François-Albert Laurent

FORCES

- Ce sont des techniques statistiques simples et faciles à interpréter.
- Elles offrent la possibilité d'avoir un modèle mixte, contenant des prédicteurs de nature continue ou catégorielle/dichotomique.
- Ce sont des techniques largement utilisées et très bien documentées.

LIMITES

- Ce sont des techniques statistiques dont la qualité des résultats est hautement dépendante de la qualité du devis employé et des données obtenues.
- Elles nécessitent le respect de nombreuses conditions pour atteindre des résultats fiables (c.-à-d. prémisses).
- Elles ne permettent pas d'inférer de relations causales entre les variables indépendantes et les variables dépendantes.

Il est difficile de clairement attribuer l'invention de la régression à une personne ; plusieurs ont contribué au test de régression tel qu'il est connu aujourd'hui. Notamment, Roger Joseph Boscovich, mathématicien, astronome, physicien, poète et philosophe croate (1755) fut le premier chercheur à calculer un coefficient de régression. Adrien-Marie Legendre, mathématicien français (1805), en tentant de déterminer les orbites de différents corps célestes, fut le premier à utiliser la méthode des moindres carrés. Ce n'est cependant que vers la fin du XIX[e] siècle que le terme « régression linéaire » fut officialisé après la publication d'un article par Francis Galton (1886), intitulé « Regression towards mediocrity in hereditary stature », qui démontrait à l'aide d'analyses de régressions linéaires que la taille d'enfants de parents anormalement grands régressait vers la moyenne. Pearson (1938, dans Stanton, 2001), un collègue de Galton, a par la suite poussé les mathématiques sous-jacentes à la régression et perfectionné, entre autres, les régressions multiples. Depuis, les tests de régression sont couramment utilisés dans plusieurs domaines, comme en économie, en médecine, en réadaptation, en géographie, en éducation et en psychologie pour n'en nommer que quelques-uns.

Tout comme Galton le faisait en 1886, les régressions linéaires et logistiques sont généralement utilisées dans le but de prédire une variable dépendante (VD) à partir d'une ou de plusieurs variables indépendantes (VI). Il est cependant important de distinguer la prédiction de la relation de causalité, puisque des relations peuvent paraître systématiques sans toutefois être liées par un lien de cause à effet. Tout comme les corrélations, les régressions permettent de qualifier (direction) et de quantifier (taille) la relation existant entre une ou plusieurs VI et une VD. Cependant, contrairement aux corrélations, elles répondent au principe de parcimonie en permettant d'éliminer les redondances de prédiction entre les VI. Autrement dit, les régressions permettent de déterminer les contributions uniques de chacune des VI à la prédiction de la VD et d'exclure du modèle de régression celles dont la contribution unique n'est plus significative.

Selon la nature de la VD étudiée, deux grandes familles de régression existent, soit les régressions linéaires et les régressions logistiques. Les régressions linéaires sont utilisées lorsque l'on souhaite prédire une VD continue (p. ex. temps de réponse, niveau d'anxiété mesuré à l'aide d'une échelle de Likert), alors que les régressions logistiques visent à prédire une variable catégorielle (p. ex., appartenance à un groupe, succès ou échec d'un traitement).

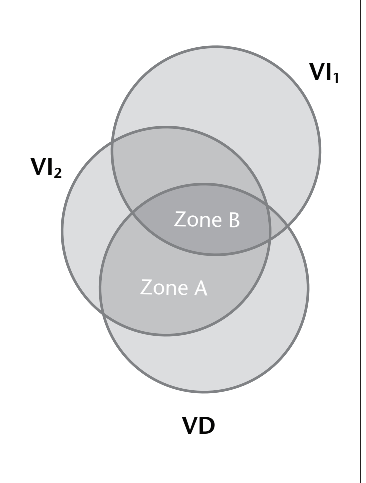
Ce chapitre est divisé en deux sections. La première porte sur les régressions linéaires, alors que la seconde traite des régressions logistiques. Chaque section comporte deux parties. Dans la première, une brève

présentation des principes mathématiques sous-jacents, des prémisses de base à respecter, des types de régressions possibles ainsi que de leurs applications sera faite. La seconde partie consiste en une illustration du test de régression. Afin de simplifier cette partie, la même banque de données a été utilisée pour l'illustration de la régression linéaire et de la régression logistique. Seule une partie de l'étude originale a été utilisée et les mêmes variables ont été conservées pour effectuer les deux analyses.

Exemple

Vous souhaitez prédire le plus efficacement possible l'intensité des symptômes du syndrome d'épuisement professionnel (*burnout*) (figure 18.1 : VD) chez des employés (Maslach et Jackson, 1981) à partir d'incidents critiques négatifs vécus en emploi (figure 18.1 : VI_1) et de la satisfaction des trois besoins psychologiques fondamentaux (figure 18.1 : VI_2) de la théorie de l'autodétermination (Deci et Ryan, 1985, 2000). Individuellement, chacune de ces VI est modérément liée à la VD. Cependant, puisque la satisfaction des trois besoins explique une proportion unique de variance (figure 18.1 : zone A), en plus de partager la proportion de variance expliquée par les incidents critiques (figure 18.1 : zone B), elle constitue à elle seule le modèle de prédiction le plus parcimonieux. Autrement dit, l'ajout de la variable *Incidents critiques* (VI_1) dans la prédiction des symptômes de *burnout* n'apporterait aucune valeur ajoutée.

Figure 18.1.
Proportion unique de variance



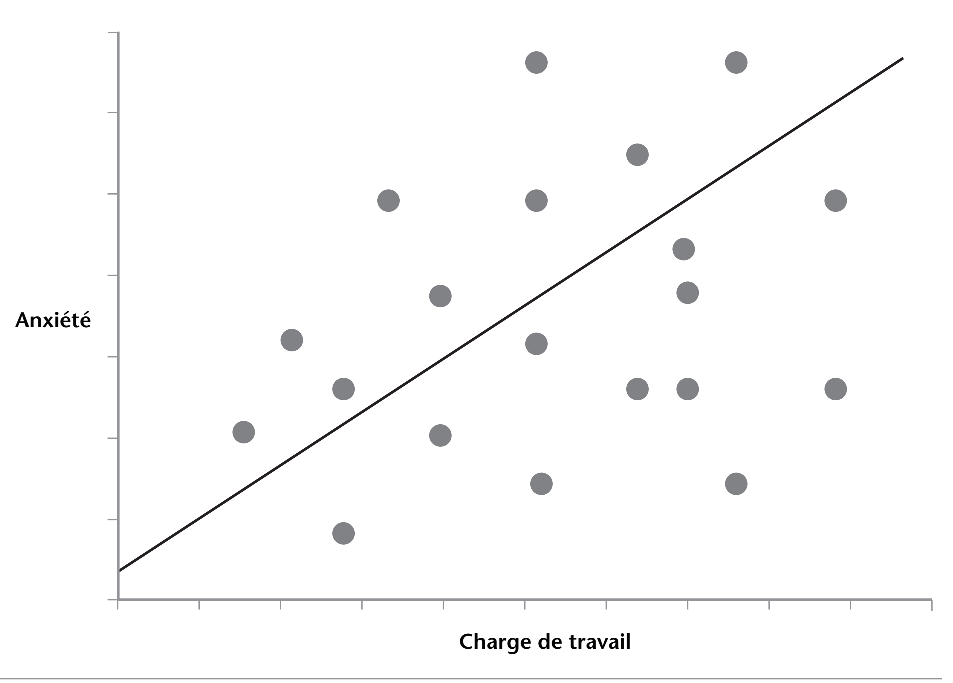
1. RÉGRESSIONS LINÉAIRES

Les régressions linéaires sont généralement utilisées dans la recherche des meilleurs prédicteurs d'un phénomène. Par exemple, elles peuvent répondre aux questions de recherche suivantes : « Quels sont les facteurs les plus importants dans la réadaptation des travailleurs à la suite d'un accident de travail ? » ou « Quelles sont les conditions de travail pouvant prédire l'anxiété des employés ? »

Essentiellement, les régressions linéaires permettent de tracer la droite qui représente le mieux l'ensemble des données observées. La droite optimale est celle qui minimise les différences (ou la distance) entre les valeurs prédites par la droite et les données observées – l'objectif est de

minimiser les valeurs résiduelles. Par exemple, nous pouvons chercher à prédire l'anxiété des enseignants à partir de la charge de travail rapportée. Chaque point de la figure 18.2 représente une observation. Le test de régression linéaire trouvera la droite qui épouse le mieux les données observées, afin d'en dégager un modèle prédictif.

Figure 18.2.
Test de régression linéaire



Équation de la régression linéaire : $\mathbf{Y = (b_0 + b_1X_1) + \varepsilon_i}$

où

- Y = variable dépendante ;
- b_0 = constante ;
- X_1 = variable indépendante ;
- b_1 = coefficient de régression ;
- ε_i = résidu.

1.1. Types de régressions linéaires

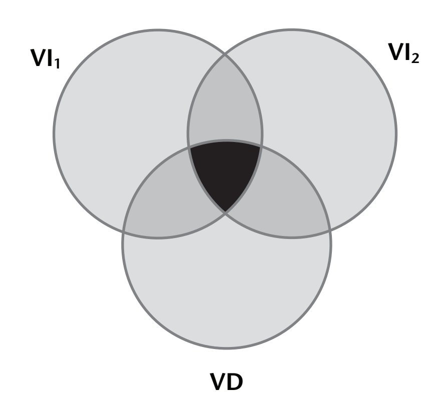
Quels prédicteurs devons-nous inclure dans notre modèle et comment les inclure ? Ces questions sont cruciales puisque le choix des variables prédictives ainsi que leur méthode d'entrée affecteront le modèle que l'on

obtiendra. Idéalement, la théorie et les résultats des études antérieures devraient guider le choix des VI. L'idée est de ne pas aller à la pêche : il ne faut pas insérer plusieurs VI avec l'espoir qu'un modèle émergera ! Une fois que le choix des VI est fait, il reste à déterminer avec quelle méthode elles seront sélectionnées dans le modèle. Trois méthodes sont communément utilisées.

1.1.1. Régression multiple standard (*standard regression/forced entry* ou *enter*)

Ce type de régression permet l'entrée simultanée de toutes les VI dans l'équation de régression. Les VI sont alors mises en compétition, sans avantage ou distinction les unes par rapport aux autres. Les aires de variance partagées sont non attribuées, ne laissant place qu'aux contributions uniques de chaque VI dans l'explication de la VD (figure 18.3).

Figure 18.3.
Régression multiple standard

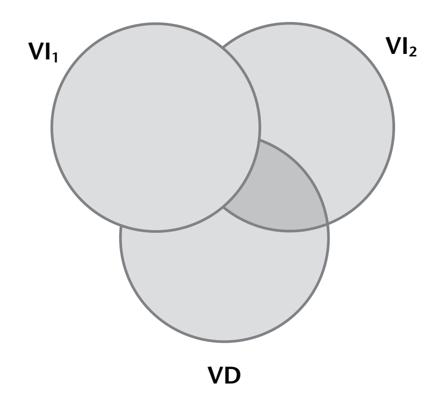


1.1.2. Régression multiple séquentielle (*hierarchical/sequential regression*)

Ce type de régression, aussi dite hiérarchique, permet une entrée séquentielle des VI dans l'équation de régression (figure 18.4). Les VI peuvent être entrées en deux ou plusieurs étapes. Habituellement, les prédicteurs connus ou les variables que nous voulons contrôler (p. ex., connaître la valeur prédictive des conditions de travail sur l'anxiété tout en contrôlant pour le sexe et l'âge du participant) constituent la première étape. Ces VI se partagent la variance de la même façon que la régression multiple standard. Les étapes subséquentes de la régression multiple hiérarchique visent à

cerner la variance unique additionnelle expliquée par les VI ajoutées : est-ce que l'ajout de nouvelles variables permet d'améliorer le modèle précédent (étape 1) en expliquant une proportion significative de la variance restante ? L'ordre d'entrée des VI est déterminé par le chercheur et est généralement sous-tendu par des fondements théoriques ou des résultats empiriques antérieurs.

Figure 18.4.
Régression multiple séquentielle



1.1.3. Régression multiple statistique (*stepwise/statistical regression*)

Ce type de régression permet l'entrée ou le retrait séquentiel de VI, une à une, de l'équation de régression sur la base d'un critère statistique. Trois méthodes de sélection sont offertes avec ce type de régression linéaire, soit l'introduction ascendante (*forward selection*), l'élimination descendante (*backward deletion*) et la régression pas à pas (*stepwise regression*). L'ordre d'entrée des VI dans l'équation est purement mathématique et est déterminée par le logiciel. Le rôle du chercheur est de repérer les VI susceptibles de prédire la VD et de les entrer en bloc dans le logiciel. Avec la *méthode d'introduction ascendante*, le logiciel inclut une première VI, celle qui contribue le plus à l'explication de la variance de la VD. Une recherche s'effectuera parmi les VI restantes afin de trouver celle qui permet d'expliquer le plus de la variance restante. Le logiciel poursuivra ainsi jusqu'à ce que le modèle ne puisse plus être amélioré de façon significative. L'*élimination descendante*, quant à elle, consiste tout d'abord à inclure toutes les VI dans l'équation et à exclure celles qui ne contribuent pas de façon significative à l'explication

de la variance de la VD. Une fois ces variables retirées, le logiciel recalculera le modèle et réévaluera la contribution des VI restantes. Encore une fois, celles qui n'atteignent pas le seuil de signification désirée seront exclues du modèle. Ce processus se poursuit jusqu'à ce que toutes les variables restantes contribuent significativement de façon unique à la variance de la VD. Finalement, la *régression pas à pas* est un compromis entre les deux précédentes méthodes de sélection. Cette méthode s'effectue en respect des principes de l'introduction ascendante, permettant cependant le retrait des VI ne contribuant plus significativement à l'explication de la variance à la suite de l'introduction d'une nouvelle VI. La technique de régression statistique peut paraître intuitive, mais elle est controversée puisque la seule logique sous-tendant le modèle de régression créé est de nature statistique. Autrement dit, les variables sont incluses et exclues du modèle sans regard à la signification théorique de celles-ci, et, par le fait même, le modèle obtenu est fondé seulement sur l'échantillon à l'étude.

1.2. Prémisses

Certaines règles doivent être respectées afin de réaliser des analyses de régressions linéaires, du moins si l'on veut interpréter correctement les résultats obtenus et les généraliser à la population à l'étude. Le tableau 18.1 présente les principales prémisses à respecter et la ou les façons de les vérifier.

Tableau 18.1.
Prémisses de la régression linéaire à respecter et leur vérification

PRÉMISSES	VÉRIFICATION
Variables à l'étude	
La VD doit être continue, alors que les VI doivent être continues ou dichotomiques. À ce titre, les variables catégorielles doivent être recodées de façon dichotomique pour permettre l'utilisation d'une telle méthode statistique (p. ex., regrouper les diagnostics psychiatriques sur la base de leur similitude en deux catégories : 1 = troubles de l'humeur, 2 = troubles anxieux).	Vérifier la nature des variables à l'étude et les convertir si nécessaire.
Une certaine proportion entre le nombre de sujets et le nombre de VI doit être respectée, mais il n'y a aucune règle unique à ce sujet. La règle proposée par Tabachnick et Fidell (2001) est de $n \geq 50 + 8m$ (où m représente le nombre de VI). Stevens (1996) recommande d'avoir au moins 15 sujets par VI. Si une régression pas à pas (*stepwise*) est effectuée, il est recommandé d'avoir au moins 40 sujets par VI.	Selon la règle adoptée, vérifier si le ratio sujet/VI est respecté. Si ce n'est pas le cas, recruter davantage de participants ou mettre de côté des VI.

(*suite*)

Tableau 18.1.
Prémisses de la régression linéaire à respecter et leur vérification (*suite*)

PRÉMISSES	VÉRIFICATION
Données extrêmes et manquantes	
Les régressions linéaires sont particulièrement sensibles aux données extrêmes uni- et multivariées, dont la présence affecte la précision des estimations. La plupart des logiciels excluent automatiquement les cas avec des données manquantes. Si plusieurs données sont manquantes, ceci affectera négativement la taille de l'échantillon. Il faut également s'assurer que les données manquantes sont aléatoires et ne forment pas de patron particulier. Ce dernier aspect est plus important que le nombre de données manquantes.	En plus d'observer les histogrammes ou les « *box-plot* » pour chacune des variables (VD et VI), il est possible de repérer les données extrêmes univariées en cernant les cas qui ont des scores standardisés (score z) élevés sur une ou plusieurs variables. Généralement, un score z plus grand ou égal à 3 ou 3,29 indique un score extrême (Seo, 2002). Les données extrêmes multivariées peuvent être repérées en calculant la distance Mahalanobis[1] et en comparant les valeurs obtenues à la valeur issue du tableau des valeurs critiques du chi-carré. Par exemple, avec trois VI et un seuil de signification de μ = 0,001, la valeur critique est de 16,266 (issue du tableau des valeurs critiques du chi-carré). La distance Mahalanobis de tous les sujets pour toutes les variables est ensuite comparée à cette valeur ; les valeurs excédant 16,266 sont considérées comme étant extrêmes. Les sujets qui ont des données extrêmes sont habituellement éliminés. L'analyse de fréquence permet de déterminer le nombre de données manquantes. Une analyse de données manquantes (SPSS MVA) permet d'établir si celles-ci sont aléatoires ou non. Habituellement, si le nombre de données manquantes est faible, soit moins de 5 %, aucun patron n'est détecté.
Multicolinéarité et singularité	
Puisque les régressions linéaires permettent d'établir les contributions uniques des VI en éliminant les redondances d'explication de variance sur la VD, le degré de corrélation entre les VI doit être pris en compte lors de leur inclusion dans le modèle de régression. L'inclusion de VI très fortement corrélées entre elles ou même indissociables statistiquement les unes des autres est à proscrire, alors que l'utilisation de VI faiblement corrélées entre elles et fortement liées à la VD est idéale pour effectuer des analyses de régressions linéaires.	Un examen de la matrice de corrélations permet un premier survol. Les corrélations entre VI de plus de 0,90 signalent un problème sur le plan de la multicolinéarité. La valeur VIF (ou la tolérance) permet aussi de vérifier la prémisse. Une valeur VIF au-delà de 10 ou une tolérance inférieure à 0,1 est problématique.

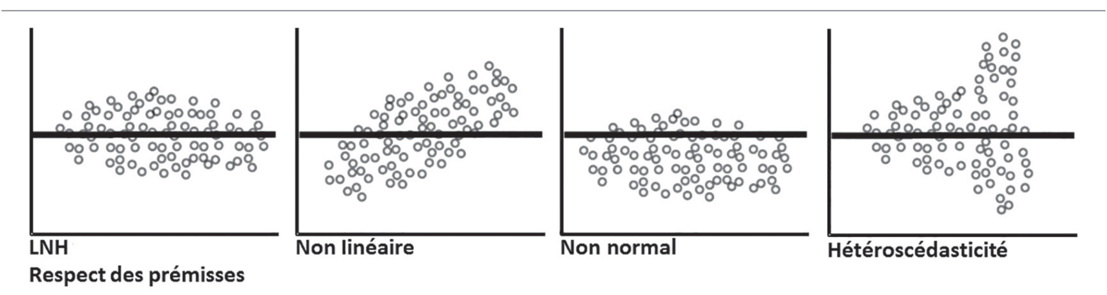
(*suite*)

Tableau 18.1. (*suite*)

Distribution des résidus

Les régressions linéaires s'appuient sur les prémisses de la normalité, de la linéarité et de l'homoscédasticité. Autrement dit, les résidus se distribuent de façon normale, expriment une relation linéaire entre les VI et la VD et démontrent que la variance expliquée de la VD est relativement la même, peu importe sa valeur.

Les graphiques de distribution des résidus[2] permettent de révéler les infractions potentielles à de telles prémisses, selon l'orientation globale qu'y prennent les résidus (voir ci-dessous).



Indépendance des erreurs

Il peut arriver que certaines erreurs de prédiction ne soient pas indépendantes. Par exemple, elles pourraient être liées à l'ordre de passation des sujets ou à certaines caractéristiques des expérimentateurs.

La valeur acceptable du test statistique Durbin-Watson se situe entre 1 et 3. Plus la valeur est près de 2, moins il y a de problèmes sur le plan de l'indépendance des erreurs.

1. Dans SPSS pour chacune des variables : *Analyze – Regression – Linear – Save – Mahalanobis.*
2. Dans SPSS : *Analyze – Descriptives – Explore – Plots – Normality plots with tests.*

1.3. Illustration de la régression multiple hiérarchique

En 2008, plus de 20 000 enseignants du Québec ont été invités à répondre à un questionnaire portant sur leur expérience de travail (Houlfort et Sauvé, 2010). De ce nombre, un peu plus de 2000 ont rempli le questionnaire. En plus d'obtenir un portrait détaillé de l'état de santé des enseignants, cette étude a permis de cerner de quelle façon les conditions de travail des enseignants affectent leur état de santé psychologique. Plus précisément, les chercheurs se basant sur la théorie de l'autodétermination (Deci et Ryan, 2000), le rôle médiateur de la satisfaction des besoins psychologiques fondamentaux entre les conditions de travail et la santé psychologique a été confirmé. La présente illustration porte sur un des objectifs secondaires de l'étude, soit de prédire l'anxiété vécue par les enseignants. Toutefois, afin de simplifier l'exemple, un sous-groupe de participants a été sélectionné (n = 358). L'âge moyen était de 41,31 ans (é.-t. : 9,53), la majorité était des

femmes (76,4%) et 87,3% avaient un poste permanent. Tous les niveaux d'enseignement étaient représentés. Parmi les VI mesurées, seulement trois VI ont été retenues aux fins de l'illustration. La VD correspond au degré d'anxiété rapporté (traité de façon continue) alors que les VI sont les conditions de travail des enseignants: le soutien de la direction, la gestion de classe et la charge de travail.

1.3.1. Vérification des prémisses

Avant de procéder à l'analyse de régression proprement dite, le respect des prémisses de base est vérifié. Dans un premier temps, chacune des observations est inspectée de façon à repérer les valeurs «hors normes». Par exemple, pour les trois VI d'intérêt, l'échelle de mesure s'échelonnait de 1 = pas du tout à 5 = toujours. L'inspection des analyses de fréquences nous permet de confirmer que toutes les valeurs respectent leur échelle de mesure.

Dans un deuxième temps, nous avons vérifié le nombre de données manquantes et si un patron pouvait être détecté parmi celles-ci. Les résultats des analyses de fréquences ont révélé très peu de données manquantes (soit moins de 5%) pour chacune des variables. En fait, seulement deux participants ont des données manquantes, ce qui exclut la possibilité d'un quelconque patron parmi celles-ci. Ce résultat, jumelé au fait que la taille de l'échantillon est relativement grand (n = 358), nous permet d'utiliser plusieurs techniques de remplacement de données (p. ex., remplacement par la moyenne, supprimer les participants avec des données manquantes, méthodes espérance-maximisation ou *expectation maximization*). Dans cet exemple, étant donné le très faible nombre de sujets avec des données manquantes nous avons opté pour l'élimination de ceux-ci. Ainsi, deux sujets ont été retranchés de l'analyse, portant l'échantillon à 356 participants.

L'analyse des diagrammes de dispersion des résidus nous informe de la *normalité*, de la *linéarité* et de l'*homoscédasticité*. Dans le présent exemple, le diagramme de dispersion des résidus ne semble pas révéler de problème sur le plan de la linéarité (l'agglomération de points ne suit pas une courbe), de la normalité, et ni de répartition non aléatoire autour de zéro (forme d'entonnoir) sous-tendant un problème d'hétéroscédasticité. Si l'une de ces prémisses étaient violées, une inspection détaillée des variables aurait été nécessaire afin de cerner et de résoudre les problèmes[1].

1. Souvent, le non-respect de ces prémisses est causé par une distribution des données qui ne respecte pas la normalité. Ainsi, l'inspection des valeurs d'asymétrie et d'aplatissement (*skewness* et *kurtosis*) ainsi que l'application des transformations nécessaires (log, racine carré, inverse, reflet racine carré, reflet log ou reflet inverse) le cas échéant, sont une première stratégie à adopter pour résoudre les problèmes détectés.

La distance Mahalanobis nous permet de rechercher les *données extrêmes* multivariées[2]. Avec trois VI et un seuil de signification de $\mu = 0{,}001$, la valeur critique est de 16,266 (issue du tableau des valeurs critiques du χ^2). Dans notre exemple, aucun cas n'excède cette valeur, ce qui indique qu'aucune donnée extrême multivariée n'est présente.

La matrice de corrélation (sortie 1 des résultats de la régression multiple) permet d'effectuer un premier survol des relations unissant la VD et les VI, en plus de procéder à un premier diagnostic de *multicolinéarité*. Les corrélations entre les VI n'indiquent pas de problème de multicolinéarité ($r < 0{,}9$). De plus, tous les indices de tolérance présentés dans la sortie 3 des résultats de la régression multiple sont supérieurs à 0,1, et tous les VIF sont inférieurs à 10. La valeur du test Durbin-Watson se situe entre 1 et 3.

Les vérifications de bases effectuées et les problèmes corrigés, nous sommes prêts à faire l'analyse de régression.

1.3.2. Régression multiple hiérarchique

SPSS REGRESSION[3] a été utilisé pour faire une régression multiple hiérarchique entre l'anxiété des enseignants et trois de leurs conditions de travail, tout en contrôlant pour l'âge et le sexe des participants. Il est ainsi recommandé d'entrer ces deux variables en premier dans le modèle (bloc 1) et ensuite, dans un deuxième bloc, les trois VI d'intérêt. La régression multiple hiérarchique permet ainsi d'évaluer la relation entre la VD et les VI, la proportion de variance de la VD expliquée par la régression et l'importance relative des VI à la solution, une fois que l'influence de l'âge et du sexe est prise en compte.

Les premiers tableaux produits par le logiciel SPSS (tableau 18.2) nous montrent pour chacune des variables du modèle la moyenne, l'écart-type et le nombre de répondants, ainsi que la matrice de corrélation[4].

2. SPSS REGRESSION permet d'obtenir la distance Mahalanobis pour chacune des VI.
3. Pour faire une analyse de régression multiple hiérarchique, il faut sélectionner, dans le menu *ANALYSE*, *RÉGRESSION* et ensuite *LINÉAIRE*.
4. Dans la sortie 1, les informations quant au sexe des participants sont inadéquates.

Tableau 18.2.
Sortie 1 des résultats de la régression multiple

Statistiques descriptives

	Moyenne	Ecart-type	N
anxiete	2,0117	,69467	356
Age	41,40	9,522	356
Sexe	1,23	,423	356
charge_travail	3,4073	1,02396	356
gestion_classe	3,0209	,83492	356
relations_direction	2,2865	1,05320	356

Corrélations

		anxiete	Age	Sexe	charge_travail	gestion_classe	relations_direction
Corrélation de Pearson	anxiete	1,000	-,029	-,025	,401	,339	,389
	Age	-,029	1,000	,165	-,038	-,023	,061
	Sexe	-,025	,165	1,000	-,090	,027	,009
	charge_travail	,401	-,038	-,090	1,000	,449	,431
	gestion_classe	,339	-,023	,027	,449	1,000	,461
	relations_direction	,389	,061	,009	,431	,461	1,000
Sig. (unilatérale)	anxiete	.	,290	,317	,000	,000	,000
	Age	,290	.	,001	,239	,336	,125
	Sexe	,317	,001	.	,046	,309	,431
	charge_travail	,000	,239	,046	.	,000	,000
	gestion_classe	,000	,336	,309	,000	.	,000
	relations_direction	,000	,125	,431	,000	,000	.
N	anxiete	356	356	356	356	356	356
	Age	356	356	356	356	356	356
	Sexe	356	356	356	356	356	356
	charge_travail	356	356	356	356	356	356
	gestion_classe	356	356	356	356	356	356
	relations_direction	356	356	356	356	356	356

La sortie 2 des résultats de la régression multiple (tableau 18.3) présente des informations cruciales quant au modèle que l'on évalue. Premièrement, il est important de noter que deux modèles sont présentés ici, puisque nous avons précisé un modèle hiérarchique à deux étapes ; une première avec nos variables contrôles, l'âge et le sexe (bloc 1), et une deuxième avec nos VI d'intérêt, soit les trois conditions de travail (bloc 2). Le modèle 1 renvoie à la première étape de la hiérarchie (âge et sexe comme prédicteurs de la VD), alors que la deuxième étape de la hiérarchie est représentée par le modèle 2, intégrant l'ensemble des variables (âge, sexe et les trois conditions de travail). Il est important de noter que le logiciel SPSS inclut automatiquement les variables du bloc 1 dans le bloc 2, même si la variance expliquée par les variables du bloc 1 s'avère non significative. Ceci permet

de vérifier la valeur ajoutée des VI d'intérêt incluses uniquement au bloc 2 : tout en contrôlant pour l'âge et le sexe, quelle est la variance expliquée par nos VI d'intérêt ?

Tableau 18.3.
Sortie 2 des résultats de la régression multiple

Récapitulatif des modèles[c]

Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation	Changement dans les statistiques				
					Variation de R-deux	Variation de F	ddl1	ddl2	Sig. Variation de F
1	,036[a]	,001	-,004	,69619	,001	,229	2	353	,795
2	,480[b]	,230	,219	,61389	,229	34,664	3	350	,000

a. Valeurs prédites : (constantes), Sexe, Age

b. Valeurs prédites : (constantes), Sexe, Age, gestion_classe, charge_travail, relations_direction

c. Variable dépendante : anxiete

ANOVA[c]

Modèle		Somme des carrés	ddl	Moyenne des carrés	D	Sig.
1	Régression	,222	2	,111	,229	,795[a]
	Résidu	171,090	353	,485		
	Total	171,312	355			
2	Régression	39,413	5	7,883	20,916	,000[b]
	Résidu	131,900	350	,377		
	Total	171,312	355			

a. Valeurs prédites : (constantes), Sexe, Age

b. Valeurs prédites : (constantes), Sexe, Age, gestion_classe, charge_travail, relations_direction

c. Variable dépendante : anxiete

Le R est le coefficient de corrélation multiple entre les VI et la VD. Le R-deux (R^2) indique le pourcentage de variance de la VD expliquée par les VI. Dans le modèle 1 (note *a*), l'âge et le sexe des participants expliquent seulement 0,001 % de la variance, alors que dans le modèle 2 (note *b*), avec les trois VI, on peut voir qu'elles expliquent 23 % de l'anxiété vécue par les enseignants. La différence de variance expliquée entre les modèles 2 et 1 (variation de R-deux) est donc de 22,9 % et elle est significative (sig. variation de F ; $p < 0,000$). Les résultats des analyses de variances (ANOVA) indiquent si l'utilisation de nos modèles est meilleure pour prédire l'anxiété des enseignants que le simple fait d'utiliser les moyennes obtenues aux VI. Le premier modèle n'est pas statistiquement significatif F (2, 353) = 0,229, $p = 0,795$, alors que le deuxième l'est, F (5, 350) = 20,916 ($p < 0,000$). Ainsi, le deuxième modèle est meilleur que le premier pour prédire la VD et il est meilleur que la simple utilisation de la moyenne comme méthode de prédiction.

Le tableau 18.4 présente les paramètres des modèles[5]. Les coefficients bêta standardisés (β; 3[e] colonne) sont suivis d'un test-*t* afin de déterminer si la contribution de la variable à la régression est significative. Ainsi, dans le premier modèle, nous pouvons voir que ni le sexe ($\beta = -0{,}021$, $p = 0{,}630$) ni l'âge ($\beta = -0{,}026$, $p = 0{,}697$) ne parviennent à prédire l'anxiété chez les enseignants. Dans le second modèle, lorsque toutes les variables sont prises en considération, la gestion de classe ($\beta = 0{,}123$, $p < 0{,}000$), la charge de travail ($\beta = 0{,}245$, $p < 0{,}000$) et la direction ($\beta = 0{,}229$, $p < 0{,}000$) contribuent significativement à la prédiction de l'anxiété vécue par l'enseignant.

Tableau 18.4.
Sortie 3 des résultats de la régression multiple

Coefficients[a]

Modèle		Coefficients non standardisés		Coefficients standardisés			Corrélations			Statistiques de colinéarité	
		A	Erreur standard	Bêta	t	Sig.	Corrélation simple	Partielle	Partie	Tolérance	VIF
1	(Constante)	2,133	,184		11,576	,000					
	Age	-,002	,004	-,026	-,482	,630	-,029	-,026	-,026	,973	1,028
	Sexe	-,034	,088	-,021	-,389	,697	-,025	-,021	-,021	,973	1,028
2	(Constante)	,891	,215		4,141	,000					
	Age	-,002	,003	-,031	-,646	,519	-,029	-,034	-,030	,964	1,037
	Sexe	-,006	,079	-,004	-,075	,941	-,025	-,004	-,003	,960	1,042
	charge_travail	,166	,037	,245	4,448	,000	,401	,231	,209	,723	1,382
	gestion_classe	,102	,046	,123	2,201	,028	,339	,117	,103	,706	1,416
	relations_direction	,151	,037	,229	4,135	,000	,389	,216	,194	,718	1,393

a. Variable dépendante : anxiete

2. RÉGRESSIONS LOGISTIQUES

L'objectif de la régression logistique est de correctement prédire une VD catégorielle à partir de VI (catégorielles ou continues). Par exemple, la régression logistique permet de répondre à des questions telles que : « Est-ce que l'accès à un traitement psychologique, pharmacologique et physiologique permet de prédire le retour ou non au travail ? », « Quelle est la probabilité de réintégrer son emploi en respectant le plan de traitement ? » et « Quelles sont les conditions de travail pouvant prédire la présence ou l'absence d'anxiété ? »

D'une certaine façon, la régression logistique est une régression linéaire, mais avec une VD catégorielle. Ainsi, ce type de régression permet de prédire l'appartenance à un groupe (p. ex., malade ou en santé) ou la

5. Seul le coefficient standardisé bêta sera discuté dans cette section. Le bêta est la version standardisée des valeurs b. Il est plus facile à interpréter puisqu'il ne dépend pas de l'unité de mesure de la variable ; il est standardisé et donc directement comparable aux autres coefficients bêta. Les coefficients bêta standardisés indiquent le changement subi par la VD (en écart-type) à chaque écart-type de changement de la VI.

probabilité qu'un phénomène se produise (p. ex., présence ou absence). La régression logistique peut être binomiale (VD avec seulement deux catégories) ou polynomiales (VD avec plus de deux catégories). Les explications qui suivent s'appliquent aux deux types de régression logistique.

2.1. Principes mathématiques généraux

Les régressions logistiques sont essentiellement régies par les mêmes principes mathématiques que les régressions linéaires. Toutefois, puisque la relation n'est pas linéaire, car la VD est catégorielle, une transformation logarithmique est nécessaire afin de rendre celle-ci linéaire.

Il est possible de constater que l'équation de la régression logistique est somme toute la même que la régression linéaire à laquelle une transformation logistique a été ajoutée. Les résultats (P(Y)) se situent entre 0 et 1.

$$P(Y) = \frac{e^{-(b_0 + b_1 X_1)}}{1 + e^{-(b_0 + b_1 X_1)}}$$

où

P(Y) = probabilité que Y arrive ;

e = base des logarithmes naturels ;

b_0 et b_1 = combinaison linéaire des prédicteurs et de la constante.

2.2. Prémisses

L'emploi des régressions logistiques, tout comme celui des régressions linéaires, dépend du respect des prémisses concernant les données extrêmes et manquantes, la multicolinéarité et la singularité des variables à l'étude, ainsi que l'indépendance des erreurs. Toutefois, d'autres prémisses doivent être respectées afin de pouvoir généraliser les résultats à la population visée (tableau 18.5).

2.3. Types de régressions logistiques

Il existe trois types de régressions logistiques : 1) directes, 2) séquentielles et 3) statistiques. Tout comme les régressions linéaires, les régressions logistiques offrent différentes stratégies d'inclusion des variables prédictrices dans le modèle de régression. Ces types de régressions logistiques (directes, séquentielles et statistiques) sont essentiellement les mêmes que pour les régressions linéaires (standards, séquentielles et statistiques).

Tableau 18.5.
Prémisses de la régression logistique à respecter et leur vérification

PRÉMISSES	VÉRIFICATION
Variables à l'étude	
La VD doit être catégorielle, alors que les VI peuvent être continues ou dichotomiques.	On suggère minimalement dix observations (p. ex., participants) par VI (Cohen, 1992).
Il doit y avoir assez de sujets relativement au nombre de variables prédictrices.	Quand une VI catégorielle est utilisée, aucune cellule ne doit avoir moins d'un cas. Un maximum de 20 % des cellules peuvent avoir cinq cas ou moins.
Les groupes ne doivent pas être parfaitement séparés. Une séparation parfaite se produit lorsque tous les cas d'un groupe ont la même valeur à une variable prédictrice (p. ex., toutes les personnes malades travaillent dans un milieu humide) alors que tous ceux d'un autre groupe ont une autre valeur sur la variable (p. ex., toutes les personnes en santé travaillent dans un milieu sec).	Des paramètres et des erreurs standardisées trop grands sont un signe qu'il y a trop peu de cas (p. ex., participants) par cellule. Des tentatives infructueuses de convergences (le logiciel ne parvient pas à une solution) sont également un signe qu'un problème existe.
Indépendance des observations et des résiduels	
Un cas (p. ex., participant) ne peut pas faire partie de deux groupes à la fois (chaque cas est unique).	
Linéarité du logarithme	
Les régressions logistiques se basent sur la prémisse que les VI sont linéairement liées à la transformation logarithmique de la VD.	La violation de cette prémisse peut être vérifiée à l'aide de l'approche Box-Tidwell (Hosmer et Lemeshow, 2000), dans laquelle les interactions entre les VI et leur transformation logarithmique sont incluses en tant que variables dans une régression logistique. Si l'une de ces interactions est significative, il y a violation de la prémisse de linéarité du logarithme ; cela peut être corrigée à l'aide de la transformation des prédicteurs fautifs. La distribution de ceux-ci (asymétrie et aplatissement) déterminera le type de transformation à faire (racine carrée, reflet inverse, etc.).

2.4. Illustration de la régression logistique directe

Dans cet exemple, nous désirons prédire l'occurrence de l'anxiété chez les enseignants à partir de trois prédicteurs : la charge de travail, les difficultés de gestion de la classe et la qualité des relations avec la direction (tableau 18.6). Aux fins de l'illustration, nous avons créé une variable dichotomique,

soit 0 = pas d'anxiété et 1 = présence d'anxiété[6]. Les trois variables indépendantes sont continues et ont été entrées en même temps dans l'équation. La linéarité du logarithme a été vérifiée à l'aide de l'approche Box-Tidwell. Pour ce faire, trois nouvelles variables ont été créées : les logarithmes de chacune des VI d'origine. Aucune des interactions n'était significative ($p < 0{,}05$), ce qui indique le respect de cette prémisse[7].

Tableau 18.6.
Sortie 1 des résultats de la régression logistique

Variables dans l'équation

		A	E.S.	Wald	ddl	Sig.	Exp(B)
Etape 1[a]	gestion_classe	-2,669	1,821	2,148	1	,143	,069
	charge_travail	,837	1,738	,232	1	,630	2,309
	relations_direction	,601	1,030	,341	1	,559	1,824
	LGC by gestion_classe	3,325	1,996	2,774	1	,096	27,805
	LRD by relations_direction	-,389	1,226	,101	1	,751	,678
	LCT by charge_travail	-,322	1,821	,031	1	,860	,725
	Constante	-1,017	3,334	,093	1	,760	,362

a. Variable(s) entrées à l'étape 1 : gestion_classe, charge_travail, relations_direction, LGC * gestion_classe , LRD * relations_direction , LCT * charge_travail .

Les premiers résultats présentés par le logiciel SPSS font référence au « bloc 0 », soit le modèle de base où seule la constante est entrée dans l'équation. Le tableau « Historique des itérations » indique que le log-vraisemblance du modèle de base est 475,18. Ceci représente l'ajustement du modèle de base, sans variable prédictrice. Ce modèle classifie correctement 66,8 % des cas (tableau 18.7). Autrement dit, si on catégorise tous les enseignants dans le groupe « absence d'anxiété » (soit la catégorie la plus fréquente), cette prédiction serait véridique dans 66,8 % des cas. Nous cherchons à améliorer ce taux en ajoutant des variables prédictrices. Le modèle de base est résumé dans le tableau « Variables dans l'équation » (tableau 18.7). Pour le moment, la seule donnée que nous avons est la valeur de la constante, soit 0,70. Le tableau « Variables hors de l'équation » dans la sortie 2 (tableau 18.7) des résultats de la régression logistique nous informe que l'inclusion d'autres variables pourrait augmenter le pouvoir prédictif du modèle. En effet, les

6. Habituellement, la VD devrait être une vraie variable dichotomique et non une variable continue divisée en deux groupes. Beaucoup de nuances nous échappent en procédant de cette façon. Mais aux fins de notre illustration, nous avons décidé de conserver les mêmes variables que celles utilisées pour l'illustration de la régression linéaire.
7. Les autres prémisses ont été examinées (données extrêmes et manquantes, multicolinéarité et singularité) dans l'exemple de la régression linaire.

trois variables indépendantes ont un score de Roa significatif (gestion de classe 29,66 ; charge de travail 39,34 ; relations avec la direction 29,39), ce qui indique que leur inclusion augmenterait de façon significative l'ajustement du modèle.

Tableau 18.7.
Sortie 2 des résultats de la régression logistique

Bloc 0 : bloc de départ

Historique des itérations[a,b,c]

Itération		-2log-vraisemblance	Coefficients
			Constant
Etape 0	1	475,245	-,674
	2	475,183	-,701
	3	475,183	-,701

a. La constante est incluse dans le modèle.

b. -2log-vraisemblance initiale : 475,183

c. L'estimation a été interrompue au numéro d'itération 3 parce que les estimations de paramètres ont changé de moins de ,001.

Tableau de classement[a,b]

Observations			Prévisions		
			anxiete		Pourcentage correct
			,00	1,00	
Etape 0	anxiete	,00	250	0	100,0
		1,00	124	0	,0
	Pourcentage global				66,8

a. La constante est incluse dans le modèle.

b. La valeur de césure est ,500

Variables dans l'équation

		A	E.S.	Wald	ddl	Sig.	Exp(B)
Etape 0	Constante	-,701	,110	40,752	1	,000	,496

Variables hors de l'équation

			Score	ddl	Sig.
Etape 0	Variables	gestion_classe	29,662	1	,000
		charge_travail	39,341	1	,000
		relations_direction	29,391	1	,000
	Statistiques globales		52,643	3	,000

Les résultats suivants portent sur le « bloc 1 », soit le modèle avec les trois variables indépendantes.

Le log-vraisemblance est de 418,98, ce qui représente une valeur inférieure au log-vraisemblance du modèle de base (475,18) : c'est ce qui est souhaité. Le tableau « Tests de spécification du modèle » dans la sortie 3 (tableau 18.8) des résultats de la régression logistique indique un chi-carré significatif de 56,20 (soit la différence entre les log-vraisemblance du bloc 0 et du bloc 1). Il est donc possible de conclure que le modèle avec les trois variables indépendantes obtient un meilleur ajustement que le modèle de base.

Tableau 18.8.
Sortie 3 des résultats de la régression logistique

Block 1 : Méthode = Entrée

Historique des itérations[a,b,c,d]

Itération		-2log-vraisemblance	Coefficients			
			Constant	gestion_classe	charge_travail	relations_direction
Etape 1	1	422,357	-3,393	,284	,384	,241
	2	419,015	-4,267	,360	,514	,269
	3	418,979	-4,370	,367	,531	,271
	4	418,979	-4,371	,367	,532	,271
	5	418,979	-4,371	,367	,532	,271

a. Méthode : Entrée

b. La constante est incluse dans le modèle.

c. -2log-vraisemblance initiale : 475,183

d. L'estimation a été interrompue au numéro d'itération 5 parce que les estimations de paramètres ont changé de moins de ,001.

(*suite*)

Tests de spécification du modèle

		Khi-Chi-deux	ddl	Sig.
Etape 1	Etape	56,204	3	,000
	Bloc	56,204	3	,000
	Modèle	56,204	3	,000

Récapitulatif des modèles

Etape	-2log-vraisemblance	R-deux de Cox & Snell	R-deux de Nagelkerke
1	418,979[a]	,140	,194

a. L'estimation a été interrompue au numéro d'itération 5 parce que les estimations de paramètres ont changé de moins de ,001.

Test de Hosmer-Lemeshow

Etape	Khi-Chi-deux	ddl	Sig.
1	9,112	8	,333

Dans la sortie 4 des résultats de la régression logistique (tableau 18.9), le tableau « Variables dans l'équation » présente les statistiques pour chacune des variables indépendantes. C'est ici que la contribution de chacune des VI au modèle est évaluée. La statistique de Wald permet de déterminer si le coefficient des variables diffère significativement de zéro ; le cas échéant, il est possible de confirmer que les VI permettent de prédire la VD[8]. Dans le présent exemple, selon la statistique de Wald, la gestion de classe ($z = 4{,}82$, $p = 0{,}03$), la charge de travail ($z = 13{,}33$, $p < 0{,}001$) et les relations avec la direction ($z = 4{,}62$, $p = 0{,}03$) permettent de prédire la VD de façon significative.

8. Le test de Wald est l'option par défaut offerte par la majorité des logiciels. Toutefois, certains chercheurs (p. ex., Menard, 1995) s'interrogent sur son efficacité et suggèrent l'utilisation d'autres tests (p. ex., Likelihood-Ratio Test).

Tableau 18.9.
Sortie 4 des résultats de la régression logistique

Tableau de classement[a]

Observations			Prévisions		
			anxiete		Pourcentage correct
			,00	1,00	
Etape 1	anxiete	,00	219	31	87,6
		1,00	77	47	37,9
	Pourcentage global				71,1

a. La valeur de césure est ,500

Variables dans l'équation

		A	E.S.	Wald	ddl	Sig.	Exp(B)	IC pour Exp(B) 95%	
								Inférieur	Supérieur
Etape 1[a]	gestion_classe	,367	,167	4,822	1	,028	1,444	1,040	2,005
	charge_travail	,532	,146	13,329	1	,000	1,702	1,279	2,263
	relations_direction	,271	,126	4,622	1	,032	1,312	1,024	1,680
	Constante	-4,371	,593	54,307	1	,000	,013		

a. Variable(s) entrées à l'étape 1 : gestion_classe, charge_travail, relations_direction.

Les dernières statistiques à vérifier sont les coefficients Exp(B) du tableau « Variables dans l'équation » dans la sortie 4 (tableau 18.9) des résultats de la régression logistique. Ils indiquent le sens de la relation (positif ou négatif) et expriment la probabilité de changement de la VD en fonction d'une unité de changement de la VI. Par exemple, à chaque nouvelle unité de la variable gestion de classe, l'enseignant présente presque une fois et demie plus de risque (1,44) de se retrouver dans le groupe vivant de l'anxiété ; à chaque nouvelle unité de la variable charge de travail, l'enseignant présente 1,70 fois plus de risque d'être dans le groupe vivant de l'anxiété. Un intervalle de confiance est donné pour chaque coefficient Exp(B). La borne inférieure de l'intervalle de confiance doit être supérieure à 1 afin de confirmer que le sens de la relation observée dans l'échantillon s'applique également à la population, ce qui est le cas pour nos trois variables d'intérêt. Si l'une de ces bornes inférieures avait été de moins de 1, le sens de la relation observée dans l'échantillon (p. ex., plus l'enseignant rapporte une charge de travail importante, plus il risque d'être dans le groupe vivant de l'anxiété) pourrait être inversé dans la population (p. ex., plus l'enseignant rapporte une charge de travail importante moins il risque d'être dans le groupe vivant de l'anxiété).

Les valeurs R^2 de Cox et Snell et de Nagelkerke (qui est une version ajustée de la mesure de Cox et Snell) évaluent à quel point le modèle final épouse bien les observations. Tout comme le R^2 de la régression linéaire, plus la valeur des R^2 de la régression logistique est élevée, mieux le modèle

est ajusté aux observations[9]. Théoriquement, les indices varient entre 0 et 1 (prédiction parfaite de la VD). Dans notre cas (voir le tableau 18.8), les deux indices reflètent une taille de l'effet faible (R^2 Cox et Snell : 0,14 ; R^2 Nagelkerke : 0,19). Le test de Hosmer-Lemeshow est non significatif, ce qui indique qu'il n'y a pas de différence significative entre les valeurs observées et les valeurs prédites pour chaque cas. Un estimé de la variance expliquée peut aussi être obtenu avec l'équation suivante :

$$R^2\text{logit} = \frac{\text{2LL original} - (-\text{2LL modèle})}{\text{2LL original}}$$

$$0{,}12 = \frac{475{,}18 - 418{,}98}{475{,}18}$$

Ainsi, le modèle final prédit 12 % de la variance de la probabilité d'être dans le groupe « anxieux ». Dans la régression logistique, il existe ainsi plusieurs mesures analogues au R^2 de la régression linéaire (R^2 de Cox et Snell, R^2 de Nagelkerke, R^2 de Hosmer-Lemeshow, p^2 de McFadden, etc.). Aucune n'est toutefois universellement préférée (Tabachnick et Fidell, 2001). Il est donc préférable de rapporter plusieurs R^2 lors de la présentation des résultats.

Le « tableau de classement » issu de la sortie 4 (voir le tableau 18.9) des résultats de la régression logistique présente la classification des cas à partir du modèle final. Dans notre modèle, 87,6 % des enseignants non anxieux sont classés correctement, mais seulement 37,9 % des enseignants anxieux sont bien classés. Au total, 71,1 % des cas sont classés correctement, comparativement à 66,8 % avec le modèle initial. Selon les résultats obtenus, cette amélioration est significative, même si la taille de l'effet est minime.

CONCLUSION

Les analyses de régressions s'avèrent fort utiles pour cerner les liens unissant les VI à une VD : identifier les facteurs (VI) qui contribuent à un phénomène (VD), évaluer la relation entre des VI et une VD tout en éliminant statistiquement l'effet d'une autre VI, comparer la force de deux ensembles de VI à prédire une VD, etc. Que ce soit avec la régression linéaire (quand la VD

9. Une distinction importante doit toutefois être faite entre le R^2 de la régression linéaire et le R^2 de la régression logistique. Alors que le premier est la mise au carré du R, ceci n'est pas le cas dans la régression logistique ; il est donc question du « pseudo-R^2 ». Voir Field (2009) et Menard (1995) pour plus de détails.

est continue) ou la régression logistique (quand la VD est dichotomique/catégorielle), les analyses de régression sont des techniques faciles à utiliser, à interpréter, et sont très bien documentées. Les techniques de régressions permettent de répondre à plusieurs questions de recherche, et leur flexibilité se prête particulièrement bien à l'analyse de phénomènes complexes. Toutefois, il est important de noter que les régressions linéaires et logistiques ne permettent pas d'établir de liens de causalité entre les VI et la VD, et que la valeur de la solution obtenue n'est égale qu'à la qualité des VI incluses au départ. Autrement dit, le chercheur doit inclure toutes les VI importantes, en fonction des études antérieures, afin d'obtenir une solution théoriquement valide. Enfin, bien que flexible, l'utilisation de ces techniques exige le respect de plusieurs prémisses.

RÉFÉRENCES

COHEN (1992). « Statistical power analysis », *Current Directions in Psychological Science*, vol. 1, p. 98-101.

DECI, E.L. et R.M. RYAN (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*, New York, Plenum.

DECI, E.L. et R.M. RYAN (2000). « The "what" and "why" of goal pursuits : Human needs and the self-determination of behavior », *Psychological Inquiry*, vol. 11, p. 227-268.

FIELD, A. (2009). *Discovering Statistics Using SPSS*, 3e éd., Thousand Oaks, Sage Publications.

GALTON, F. (1886). « Regression towards mediocrity in hereditary stature », *Journal of the Anthropological Institute*, vol. 15, p. 246-263.

HOSMER, D.W. et S. LEMESHOW (2000). « Model-building strategies and methods for logistic regression », dans *Applied Logistic Regression*, 2e éd., New York, John Wiley & Sons, p. 91-142.

HOULFORT, N. et F. SAUVÉ (2010). *La santé psychologique des enseignants de la Fédération autonome de l'enseignement*, Montréal, École nationale d'administration publique.

MASLACH, C. et S.E. JACKSON (1981). « The measurement of experienced burnout », *Journal of Organizational Behavior*, vol. 2, nº 2, p. 99-113.

MENARD, S. (1995). *Applied Logistic Regression Analysis*, Thousand Oaks, Sage Publications.

PEARSON, E.S. (1938). *Mathematical Statistics and Data Analysis*, 2e éd., Belmont, Duxbury.

SEO, S. (2002). *A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets*, Mémoire de maîtrise, <http://www.d-scholarship.pitt.edu/7948/1/Seo.pdf>.

STANTON, J.M. (2001). « Galton, Pearson, and the Peas : A brief history of linear regression for statistics instructors », *Journal of Statistics Education*, vol. 9, n° 3, <http://www.amstat.org/publications/jse/v9n3/stanton.html>.

STEVENS, J. (1996). *Applied Multivariate Statistics for the Social Sciences*, 3e éd., Mahwah, Lawrence Erlbaum.

TABACHNICK, B.G. et L.S. FIDELL (2001). *Using Multivariate Statistics*, Boston, Allyn and Bacon.

CHAPITRE 19

L'ANALYSE DE SURVIE
Un exemple appliqué au maintien en emploi en entreprises sociales

Nathalie Lanctôt
Djamal Berbiche

FORCES

- Elle permet de tenir compte de la contribution du temps de suivi de chaque participant, que l'événement d'intérêt ait eu lieu ou non.
- Comme les modèles de survie exploitent autant la probabilité de l'apparition de l'événement que sa durée de survie, ceci accroît la qualité de l'inférence statistique pour les données quantitatives.
- Elle permet l'utilisation des données des participants qui ont abandonné l'étude avant la fin.

LIMITES

- L'hypothèse des risques proportionnels, qui est importante dans le modèle de Cox, nécessite que chaque caractéristique soit constante au cours du temps.
- Les valeurs extrêmes peuvent l'influencer significativement, ce qui exige de prendre une décision *a priori* sur la manière dont elles seront traitées.
- Elle exige de connaître le moment initial de l'étude et la date d'occurrence de l'événement étudié.

L'analyse des données de survie est un ensemble de méthodes statistiques qui étudie la modélisation du temps jusqu'à l'apparition ou non d'un événement d'intérêt, communément appelé «décès». Cet événement peut être positif comme une guérison ou négatif comme une perte d'emploi. D'un point de vue statistique, la nature de l'événement est certes secondaire. Elle n'étudie donc que les variables aléatoires positives puisque ce sont des durées de temps et qu'il ne peut ainsi y avoir de valeur de temps négative. Ainsi, l'analyse de survie permet de savoir non seulement si un participant a «vécu» l'événement étudié, mais surtout à quel moment cet événement est survenu. Afin de faciliter la compréhension et la lecture de ce chapitre, la perte d'un emploi sera choisie comme l'exemple d'événement d'intérêt.

Historiquement, ce type d'analyse a d'abord été utilisé en démographie au XVIIe siècle pour mesurer les espérances de vie à partir des statistiques de décès. Par la suite, avec le développement de la théorie des probabilités, les modélisations des probabilités de décès à un certain âge sont apparues, telles que les fonctions de risque. Toutefois, un essor de l'analyse de survie a eu lieu au cours des dernières décennies, notamment à la suite de la publication des résultats des modèles semi-paramétriques de Kaplan et Meier (1958) ainsi que du développement du modèle de régression semi-paramétrique de Cox (1972), qui permet de prendre en compte l'effet de plusieurs variables et donc de relier le risque étudié à plusieurs facteurs pronostiques. Ce développement était salutaire, vu que les modèles paramétriques classiques étaient peu adaptés aux modèles temporels, qui sont très souvent asymétriques. L'utilisation de l'analyse de survie en cancérologie, où la survie des patients est un élément clé pour juger de l'efficacité du traitement, a ouvert la porte à d'autres domaines scientifiques, en généralisant cette notion à tout événement de nature dichotomique. Ces méthodes sont particulièrement adaptées pour les études de cohorte, et la terminologie provient donc de ce contexte épidémiologique. Toutefois, il est important de préciser que ces méthodes se prêtent également très bien à l'analyse de problématiques reliées à la santé mentale, comme nous allons le voir dans le présent chapitre.

Ce chapitre sera divisé en deux grandes parties. La première partie théorique abordera les éléments suivants: les particularités du modèle, la terminologie, les différentes méthodes, le modèle de régression de Cox, et la taille de l'échantillon. La seconde partie consistera en un exemple appliqué de l'analyse de survie, c'est-à-dire le maintien en emploi des personnes ayant des troubles mentaux graves travaillant dans des entreprises sociales.

1. ASPECTS THÉORIQUES DE L'ANALYSE DE SURVIE

1.1. Particularités du modèle

Les données de survie sont caractérisées par plusieurs particularités que l'on ne retrouve pas toujours dans les autres modèles de régression. Ainsi, l'analyse de survie possède sa propre terminologie, riche de plusieurs concepts où la définition de chacun est indispensable pour décrire et analyser ce type de données. En effet, il apparaît primordial de commencer par définir chacun de ces termes avant de décrire les différentes étapes de l'analyse de survie.

Premièrement, dans l'analyse de survie, la variable dépendante fait référence à deux types de données recueillies : *a*) une variable dite qualitative, qui représente la survenue ou non de l'événement étudié (p. ex., la perte d'un emploi) ; et *b*) une variable dite quantitative, qui représente la durée de survie (T) sous le format continu (p. ex., le nombre de jours où la personne s'est maintenue en emploi). Dans la plupart des cas, cette durée peut être en minutes, en jours, en mois, voire en années. Il est à noter cependant qu'il est possible de considérer des événements simultanés (*ties*) tels que la perte d'un emploi dans la même journée pour deux employés.

Deuxièmement, il est important de spécifier différentes « dates » lors d'une analyse de survie. Ainsi, la date d'entrée dans l'étude du participant doit être spécifiée. Cette date est propre à chaque individu qui a répondu de façon satisfaisante aux critères d'inclusion. L'analyse de survie permet de tenir compte de différentes dates d'entrée dans l'étude. Au cours du temps, certains individus vont être soumis à l'événement étudié, comme la perte d'un emploi, le décès, la récidive, ou l'hospitalisation. Les individus qui ne seront pas soumis à l'événement au cours de l'étude (p. ex., l'étude s'arrête alors que l'individu est toujours en emploi ou le participant déménage avant la fin de l'étude) seront considérés comme étant des « données censurées ». Ces données censurées seront utilisées dans les analyses de survie, ce qui permet de maximiser toutes les informations recueillies au cours d'une étude. Voici maintenant la description des différents concepts clés de l'analyse de survie.

- *Date d'origine* : Cette date index correspond à la date d'entrée du participant dans l'étude. Cette date de référence sera éventuellement différente, car elle peut varier d'une personne à une autre. Par exemple, la date à laquelle la visite chez le psychologue a eu lieu peut être différente pour chaque participant. Ainsi, il s'agit d'une date de référence ou d'une date index à partir de laquelle on mesure le temps de survie.

- *Date de point* : On appelle date de point la date à laquelle se termine l'étude. Cette date est indépendante de la date des dernières nouvelles des participants.
- *Date des dernières nouvelles* : On s'intéresse à cette date afin de pouvoir calculer la durée de survie, qui représente la différence entre cette date de fin et la date d'origine. Cette date de fin sera l'une des dates suivantes :
  - la date de point si pendant toute l'étude le participant n'a pas encore vécu l'événement (p. ex., le participant se maintient en emploi et n'a toujours pas perdu son emploi) ;
  - la date de la censure, qui est la date de la dernière information recueillie auprès de la personne avant qu'elle (se) soit retirée de l'étude. Dans ce cas de figure, l'événement d'intérêt n'a toutefois pas encore eu lieu ;
  - La date où a lieu l'événement d'intérêt (p. ex., la date où le participant perd son emploi).
- *Temps de survie* : Le délai entre la date d'origine d'un participant et la date des dernières nouvelles s'appelle le temps de participation ou le temps de survie du participant (p. ex., le temps du maintien en emploi).
- *Recul* : Le délai entre la date d'origine d'un participant et la date de point s'appelle le recul.
- *Données censurées à droite* : Lorsque la date d'origine est connue et survient après ou juste au début de l'étude et que la date de censure est avant la date de point, alors ces données sont dites « censurées à droite ». Ceci signifie que soit *a*) la personne s'est retirée de l'étude sans avoir subi l'événement avant la fin de l'étude ; soit *b*) la personne est demeurée dans l'étude, mais n'a toujours pas subi l'événement à la fin de l'étude. Les situations *a*) et *b*) représentent la majorité des données d'analyse de survie.
- *Données censurées à gauche* : Ce sont les données qui ne nous permettent pas de connaître le temps de survie exact, car il n'est pas possible de connaître la date d'origine. La censure à gauche est moins souvent abordée dans les études épidémiologiques.
- *Fonction du risque* : Il s'agit du risque instantané de survenue de l'événement en tenant compte du temps. Cette fonction $h(t)$ correspond au taux de « mortalité » (soit le taux d'occurrence de l'événement d'intérêt) au temps t.

- *Fonction de survie S(*t*)* : La fonction de survie S(*t*) donne la probabilité qu'une personne survive au-delà d'un certain temps *t*. Elle représente également la probabilité qu'un participant soit encore « vivant » après un délai *t*, ou plus généralement la proportion de survivants après un délai *t*. La fonction de survie s'écrit de la manière suivante :

$$S(t) = P(T > t)$$

où

P = probabilité ;

T = variable aléatoire du temps de survie.

Un exemple de la fonction de survie est illustré à la figure 19.1. Plus précisément, cette figure présente en fonction du temps la proportion des personnes initialement incluses dans l'étude et qui sont toujours vivantes au temps *t*. Autrement dit, elle désigne la probabilité de survivre au moins jusqu'à la date *t*. Au début, soit au temps 0, toutes les personnes sont présentes, donc S(0) = 1, car c'est la proportion des personnes vivantes au temps 0. Si l'on suivait ces personnes pendant un laps de temps très long (∞), elles seraient toutes décédées et donc S(∞) = 0.

En général, lorsque l'on recueille des données, il est important de garder les dates de tous les événements qui se produisent et leur nature afin de pouvoir calculer adéquatement les durées de survie et de les caractériser comme étant des événements si l'événement a eu lieu ou des censures si l'événement n'a pas eu lieu.

Afin d'illustrer divers exemples de données de survie, la figure 19.2 présente cinq participants qui ont pris part à une étude qui portait sur le maintien en emploi. Sur la ligne verticale est représenté chaque participant type, alors que sur la ligne horizontale est représenté le temps en nombre de mois. Ainsi, un « X » signifie que le participant a subi l'événement (c.-à-d. qu'il a perdu son emploi). De façon plus précise, on note que dans cet exemple, le participant « retiré de l'étude » est entré dans l'étude un mois et demi après le début de l'étude, et s'est retiré six mois plus tard. Le temps de survie pour le participant « retiré de l'étude » est donc censuré à 4,5 mois (c.-à-d. que la personne s'est maintenue en emploi pendant quatre mois et demi sans avoir vécu l'événement d'intérêt, soit la perte de l'emploi). Le participant « échec » est suivi du début de l'étude jusqu'à la survenue de l'événement. Le temps de survie de ce participant est donc de 4,8 mois. Autrement dit, la personne s'est maintenue en emploi pendant plus de 4 mois avant de perdre son emploi. Il est important de noter qu'un même temps de survie peut avoir une signification différente. Ainsi, même si deux

individus ont à peu près le même temps de survie, ils vont être traités différemment étant donné que l'un s'est retiré de l'étude et que l'autre a subi l'événement d'intérêt. En effet, l'analyse de survie permet de tenir compte des données recueillies d'un participant qui abandonne l'étude en cours de route.

Figure 19.1.
Fonction de survie

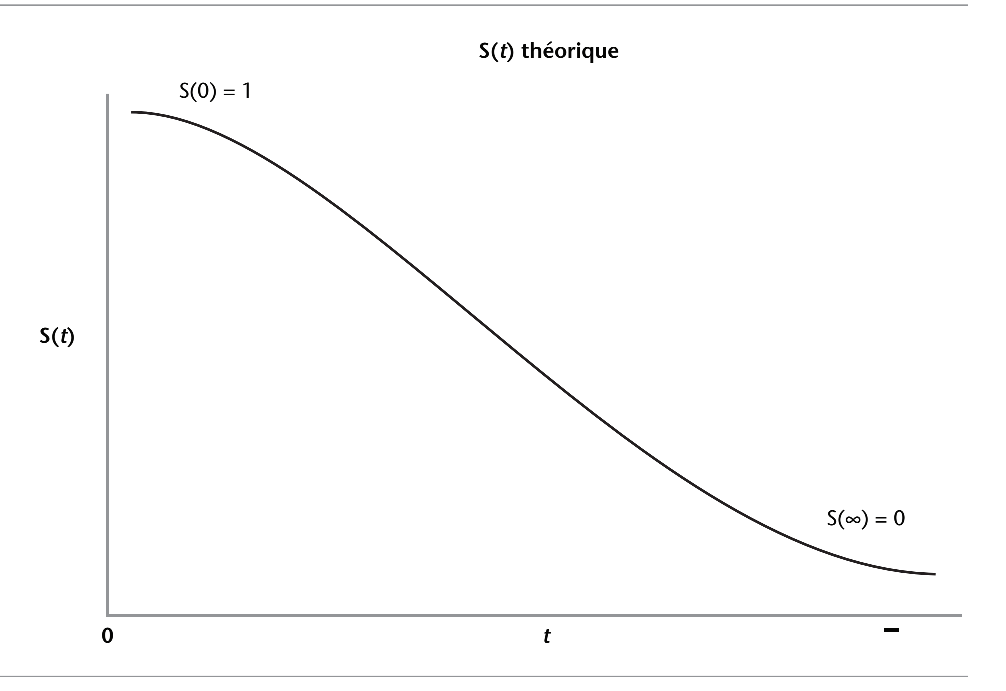
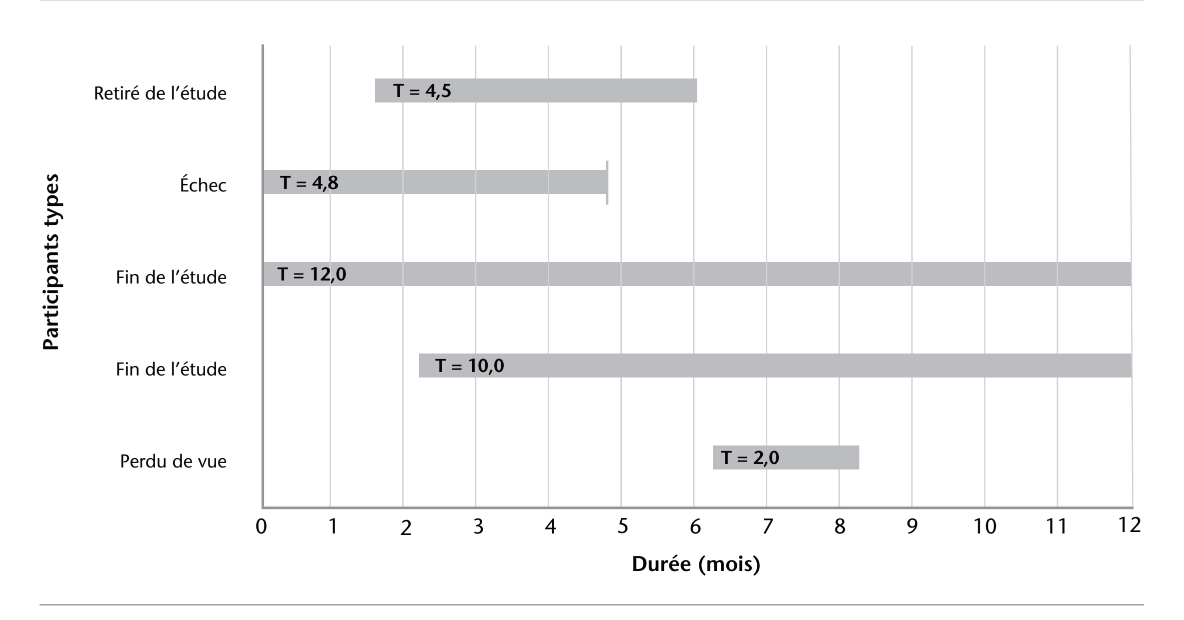


Figure 19.2.
Illustration des temps de survie



Dans la figure 19.2, l'abscisse représente la durée, soit le nombre de mois, et l'ordonnée représente les participants types.

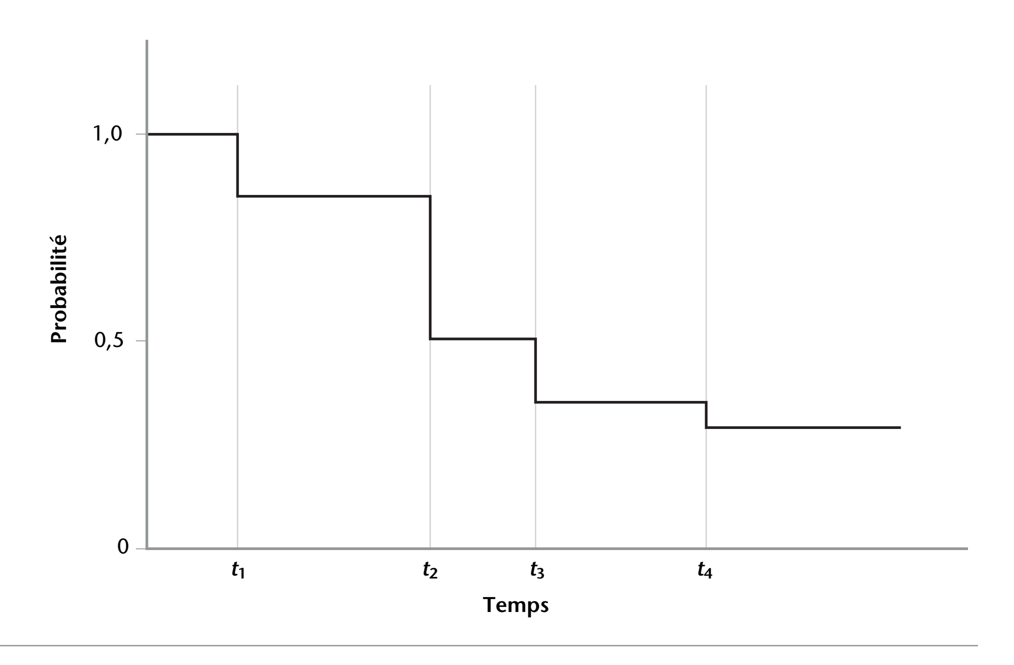
1.2. Méthodes

Il existe deux méthodes pour représenter l'évolution d'une courbe de survie : *a*) la méthode de Kaplan-Meier ; et *b*) la méthode actuarielle.

1.2.1. La méthode de Kaplan-Meier (1958)

Cette méthode est la plus utilisée. Les taux de survie conditionnels sont calculés lors de chaque survenue d'un épisode de l'événement étudié en ne gardant comme dénominateur que les participants encore « vivants » (ou qui n'ont pas vécu l'événement d'intérêt) au début de chaque intervalle. On obtient ainsi des « marches » de longueurs inégales (figure 19.3). Cette méthode est également connue sous le nom de Product Limit Estimations (PLE).

Figure 19.3.
Méthode de Kaplan-Meier (1958)

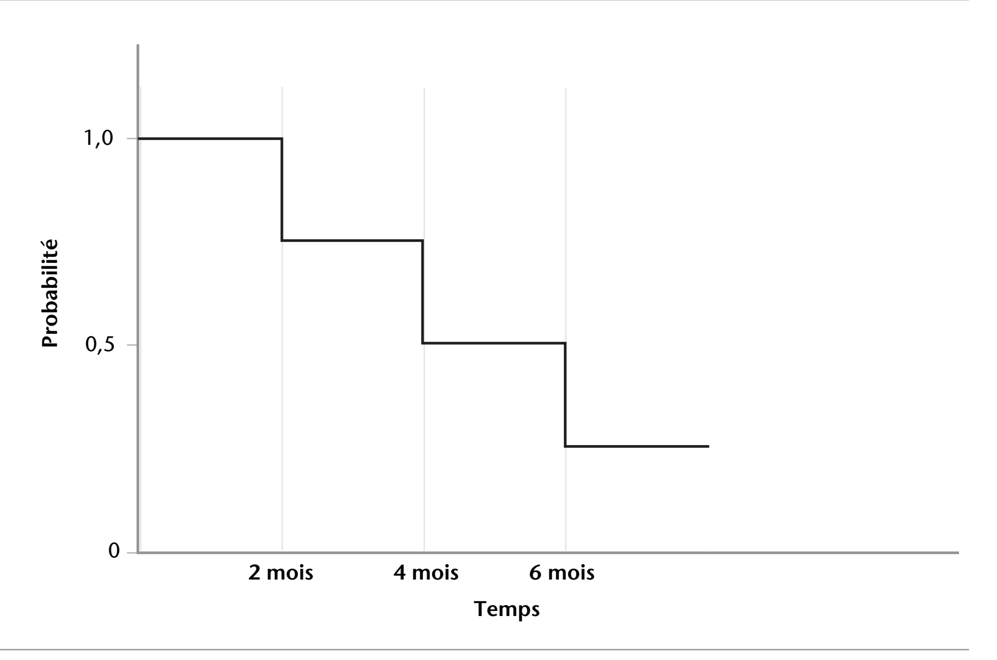


1.2.2. La méthode actuarielle (Böhmer, 1912)

Dans le cadre de cette méthode, les périodes de calcul du taux de survie sont prédéterminées et ont les mêmes intervalles en général (p. ex., 1 mois, 2 mois, 3 mois, etc.). Ainsi, les «marches d'escalier» déterminées par cette méthode sont toutes de longueur égale (figure 19.4). Le calcul de la survie se fait au milieu de chaque classe. Cette méthode tend donc à limiter le nombre de points.

Il est important de remarquer que comme toutes les durées de survie ne sont pas connues à cause des censures, l'estimation de la durée de survie moyenne n'est pas toujours possible. La médiane de survie est la mesure la plus appropriée dans les analyses de survie: elle correspond au temps où l'on observe 50% des événements des participants inclus dans l'étude. Elle doit être rapportée avec son intervalle de confiance afin de bien voir l'ampleur de l'imprécision des estimations.

Figure 19.4.
Méthode actuarielle (Böhmer, 1912)



1.3. Régression de Cox

Le modèle de régression à risques proportionnels a été initialement proposé par le statisticien britannique David Cox (1972). Ce modèle a pour but de mesurer l'effet d'un ensemble de covariables sur le temps d'apparition d'un événement en présence de censures. Il est le modèle le plus utilisé pour l'analyse des données de survie. Il s'applique à toute situation où l'on analyse le délai de survenue d'un événement. Par exemple, ce modèle peut être utilisé si l'on veut estimer l'effet de l'âge, du sexe, du revenu, des comorbidités, et le fait de suivre un programme de formation sur la survenue de l'événement d'intérêt (p. ex., la perte d'un emploi). Ceci permettra de répondre à deux types de questions : « L'un des deux programmes de formation est-il plus efficace en termes d'amélioration de la survie des personnes (c.-à-d. le maintien en emploi) ? » et « Y a-t-il d'autres facteurs qui contribuent à l'amélioration ou à la détérioration de cette survie (tels que la qualité de vie générale, l'estime de soi) ? »

Ainsi, la régression de Cox permet d'intégrer des variables explicatives pour l'estimation de la fonction de survie. Les variables indépendantes peuvent être dichotomiques (valeurs 0 et 1; p. ex., pour la variable sexe: homme ou femme), catégorielles (minimum trois catégories; p. ex., pour le revenu: < 15 000 $, de 15 000 à 20 000 $, et plus de 20 000 $) ou continues (p. ex., l'âge).

Dans la régression de Cox, ce qui est estimé, c'est le risque instantané d'occurrence de l'événement au cours de la durée de suivi. On le note $h(t)$, et il représente la probabilité (par unité de temps), pour une personne à une date donnée, de «perdre son emploi» dans l'instant qui suit. Si par exemple t est mesurée en semaines, alors $h(t)$ estimera la probabilité d'un individu travaillant à la semaine t de perdre son emploi la semaine suivante.

1.3.1. Taille de l'échantillon

Généralement, le calcul de la taille de l'échantillon dans les analyses de survie est relié au nombre d'événements que l'on pense qui vont survenir dans notre modèle. Toutefois, selon Eliason (1993), dans le cadre d'une analyse de survie, lorsque cinq variables indépendantes sont utilisées, un échantillon de 60 participants est suffisant. Ainsi, un ratio d'une variable pour dix participants est nécessaire.

1.3.2. Équation de régression

Le modèle de régression de Cox relie k variables explicatives ($X_1, \ldots, X_k$) à la fonction de risque instantané par la formule suivante, qui est le produit d'une fonction qui dépend du temps et d'une fonction ne dépendant que des variables explicatives:

$$h(t) = h_0(t)\exp(\beta_1 X_1 + \ldots + \beta_k X_k)$$

On a donc une relation log-linéaire entre la fonction de risque instantané et les covariables.

$$\mathrm{Log}(h(t)/h_0(t)) = \beta_1 X_1 + \ldots + \beta_k X_k,$$

car cette fonction Log est une fonction linéaire des X_k

Comme avec les autres méthodes multivariées, tester l'association entre les covariables X_i et la survie revient à tester si $\beta_i = 0$, $i = 1, \ldots, k$. Le terme $h_0(t)$, analogue au terme d'intercept (ou ordonnée à l'origine) dans un modèle logistique ou linéaire (voir le chapitre 18 de cet ouvrage), est appelé risque instantané de base dans l'analyse de survie. Si tous les X_k sont nuls, alors $h(t) = h_0(t)\exp(0) = h_0(t)$, qui correspond ici à un risque instantané de perdre son emploi quand toutes les autres variables sont nulles.

Supposons que dans une analyse de survie, on s'intéresse aux déterminants de la perte d'emploi d'une personne ayant un trouble mental et que l'un des déterminants est la région d'habitation notée X_1, où $X_1 = 1$ signifie *urbain* et $X_1 = 0$ signifie *rural*, alors le rapport des risques instantanés est simplement $e^{ß1}$:

$$risque = \frac{h_0(t)\ exp(\beta_1 \times 1 + \beta_2 X_2 + \ldots + \beta_k X_k)}{h_0(t)\ exp(\beta_1 \times 0 + \beta_2 X_2 + \ldots + \beta_k X_k)} = \exp(\beta_1)$$

On remarque que le risque de base n'apparaît plus dans le résultat final. Il est courant d'estimer les coefficients β sans faire d'hypothèse sur la forme de la distribution de base h_0 bien qu'il soit néanmoins possible de l'estimer. On remarque ainsi que le risque est exactement égal à $\text{Exp}(\beta_1)$. Ce dernier ne dépend donc pas de $h_0(t)$.

1.3.3. Validation du modèle de régression de Cox

Vérification de l'hypothèse des risques proportionnels

L'une des hypothèses importantes du modèle de Cox est la proportionnalité des risques. Cette hypothèse stipule que les risques sont constants et proportionnels dans le temps. Il faut donc toujours vérifier si l'hypothèse des risques proportionnels est valide. En effet, si cette hypothèse n'est pas valide, alors le modèle ne sera pas valide. Autrement dit, le ratio des risques entre deux personnes doit toujours être le même, et ce, pour tous les points dans le temps. Ainsi, pour chaque covariable introduite dans le modèle, on teste si son effet est indépendant du temps et constant (qu'il soit positif, négatif ou sans effet). Pour chaque covariable, la survie $S(t)$ est estimée dans les différents groupes de participants par la méthode de Kaplan-Meier. Si les risques sont proportionnels, les tracés correspondant aux différents groupes sont « parallèles », c'est-à-dire qu'ils ne se coupent pas ; l'hypothèse des risques proportionnels est alors valide.

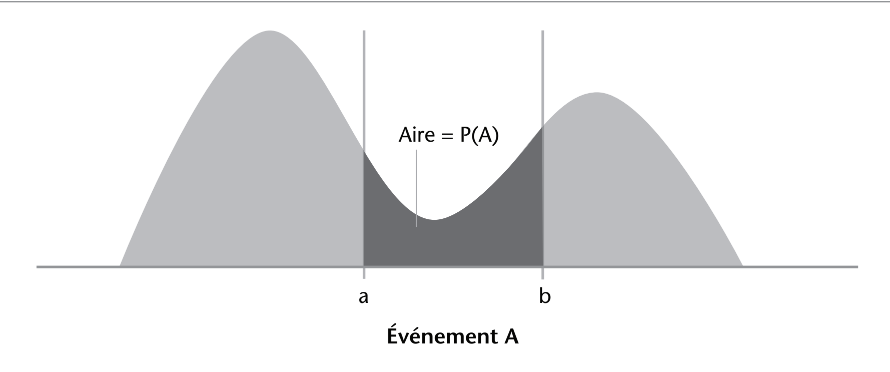
Une autre façon de tester l'hypothèse des risques proportionnels est d'inclure un terme d'interaction. Cette méthode est plus efficace et tient compte de toutes les variables, qu'elles soient continues ou catégorielles.

Types de méthode

Étant donné que la fonction de risque est supposée être « continue » dans le modèle de Cox, théoriquement, deux ou plusieurs individus ne peuvent avoir le même événement en même temps. En effet, la probabilité qu'une variable continue soit égale à une constante est toujours nulle. La probabilité

est définie comme l'aire sous la courbe entre deux valeurs (figure 19.5). Toutefois, lorsque deux valeurs sont égales, il est impossible de connaître l'aire sous la courbe.

Figure 19.5.
Aire sous la courbe entre deux valeurs



Mais comme les unités de mesure en général sont en jours, mois, années, etc., ces cas problématiques peuvent se produire (p. ex., deux individus qui perdent leur emploi la même journée). Des options existent pour répondre à ce genre de problématique. Il existe quatre méthodes pour tenir compte des différences de la survenue des temps de survie : 1) Breslow, 2) Efron, 3) Exacte, et 4) Discrète.

- La *méthode de Breslow* (Breslow, 1974) tient compte de la présence d'événements multiples à une même date. C'est la méthode par défaut utilisée par SPSS.
- La *méthode d'Efron* (Efron, 1977) est meilleure, car elle ajuste le risque en tenant compte des moyennes des individus ayant eu le même événement la même date. Allison (1995) recommande d'utiliser l'approximation d'Efron, car elle est beaucoup plus rapide que la méthode Exacte et tend à produire des estimations beaucoup plus précises que l'approche de Breslow.
- La *méthode Exacte* (Delong, Guirguis et So, 1994 ; Kalblfleisch et Prentice, 1980) suppose que les événements enregistrés la même date ont eu lieu « réellement » à des dates différentes (si nous avions pu raffiner la mesure de ces données). Autrement dit, la méthode consiste à admettre que les événements se produisent les uns à la suite des autres, et ce, sans savoir l'ordre de ceux-ci. Étant donné que l'on

ne connaît pas l'ordre exact, le nombre de possibilités peut s'avérer être très grand. Plus généralement, s'il y a *k* individus parmi *n* qui affichent le même temps d'événement, le calcul de la contribution fait intervenir *k* termes. Par exemple, lorsque seulement 10 événements ont lieu en même temps, le nombre d'évaluations possible est de 3 628 800. Ceci démontre qu'étant donné que le nombre d'événements peut être très élevé, l'analyse peut prendre un temps considérable d'exécution, car elle prend en considération toutes les possibilités.

Finalement, la *méthode Discrète* présume que les temps de survie sont discrets et qu'ils ont eu lieu exactement en même temps. Elle peut être appropriée quand les temps de survie sont regroupés en classe et qu'ils sont donc utilisés comme des données d'une distribution discrète.

Il est important de préciser que lorsque tous les événements surviennent à des dates différentes, les quatre méthodes deviennent équivalentes, c'est-à-dire qu'elles donnent sensiblement les mêmes résultats.

1.3.4. Coefficient, interprétation et logiciels

Afin de bien comprendre les résultats de l'analyse de régression de Cox, il faut prendre note que la valeur Exp(β) représente la valeur pour laquelle le ratio des fonctions du risque augmente ou diminue lorsque *x* varie d'une unité. Il s'agit en fait du ratio des risques (RR), intitulé «*hazard ratio*» en anglais[1]. Si le RR est égal à 1, alors le risque d'avoir l'événement d'intérêt (p. ex., la perte d'emploi) au temps *t* reste le même lorsque *x* varie d'une unité (donc ce facteur n'a pas d'influence). Si le RR est inférieur à 1, alors le risque de perdre son emploi au temps *t* diminue lorsque *x* augmente d'une unité. Finalement, si le RR est supérieur à 1, alors le risque de perdre son emploi au temps *t* augmente lorsque *x* augmente d'une unité. Le RR est généralement présenté avec son intervalle de confiance avec une borne minimale strictement positive et une borne maximale. Ainsi, si la valeur 1 est comprise dans cet intervalle de confiance, alors la variable indépendante correspondante n'est pas statistiquement significative au seuil $\alpha = 0{,}05$ par défaut. Par exemple, si l'on mesure de façon catégorielle la satisfaction au travail (où 1 signifie que la personne est satisfaite et 0, que

1. Différence entre le rapport de risque instantané (*hazard ratio*) et le rapport de cotes (*odds ratio*) : un rapport de cotes est un rapport de risques entre un groupe expérimental et un groupe témoin. Le risque instantané est la probabilité de l'événement par unité de temps. Le rapport de risque instantané est le rapport du risque instantané entre les exposés et les non-exposés.

la personne est non satisfaite) et que le risque est de 1,46, ceci signifie que les personnes non satisfaites ont 46 % plus de chances de perdre leur emploi que les personnes satisfaites[2].

Les logiciels SPSS (Statistical Package for Social Sciences) et SAS (Statistical Analysis System) analysent facilement les données de survie. Les logiciels Stata, Statistica et S-PLUS peuvent aussi être utilisés pour faire ce type d'analyse.

2. UN EXEMPLE D'APPLICATION DE L'ANALYSE DE SURVIE POUR L'ÉTUDE DU MAINTIEN EN EMPLOI DANS LES ENTREPRISES SOCIALES

2.1. Mise en contexte

Le travail s'avère très important dans le processus de réadaptation des personnes qui présentent un trouble mental. Toutefois, malgré l'importance que le travail peut avoir dans le processus de rétablissement, seulement 20 % des personnes ayant un trouble mental occupent un emploi. Qui plus est, une fois que les personnes ont obtenu un emploi, le maintien au travail reste une problématique complexe. En effet, des études soulignent la courte durée du maintien en emploi sur le marché régulier du travail pour les personnes avec des troubles mentaux (Corbière *et al.*, 2006 ; Xie *et al.*, 1997). Parmi les facteurs qui ont été identifiés dans la littérature scientifique comme jouant un rôle important dans le maintien en emploi, notons l'âge, la présence de symptômes psychiatriques, l'estime de soi, la satisfaction au travail, la qualité de vie générale et la qualité de vie au travail (Bassett, Lloyd et Bassett, 2001 ; Catty *et al.*, 2008 ; Corbière *et al.*, 2006 ; Mueser, Salyers et Mueser, 2001 ; Resnick et Bond, 2001 ; Werwiorski et Fabian, 2004).

Il existe une alternative au marché du travail ordinaire pour les personnes avec des troubles mentaux : les entreprises sociales (Dumais et Vaillancourt, 2005). Une entreprise est dite sociale lorsqu'elle prône une mission sociale qui consiste à intégrer en emploi des personnes ayant des troubles mentaux graves ou présentant d'autres difficultés majeures d'insertion socioprofessionnelle, qu'elle est privée et, enfin, organisée à partir d'une démarche entrepreneuriale. En effet, ces entreprises offrent des avantages

2. Pour avoir un exemple d'une variable ayant un risque inférieur à 1, le lecteur est invité à prendre connaissance de l'illustration dans la seconde partie de ce chapitre.

qui favoriseraient un plus long maintien en emploi, mais peu d'études ont évalué le maintien en emploi chez les personnes avec un trouble mental et œuvrant dans des entreprises sociales (Corbière, Lecomte et Lanctôt, 2011).

2.2. Objectif de l'étude

L'objectif de cette étude consistait donc à évaluer les déterminants du maintien en emploi des personnes avec un trouble mental travaillant dans des entreprises sociales québécoises. Pour ce faire, une étude longitudinale a été conduite (Lanctôt, Corbière et Durand, 2012).

2.3. Méthodologie

2.3.1. Participants

La population cible de l'étude était des personnes avec un trouble mental et travaillant dans des entreprises sociales québécoises. Les participants devaient répondre aux critères de sélection suivants : 1) avoir un diagnostic de trouble mental (p. ex., schizophrénie) ; 2) travailler dans une entreprise sociale ; et 3) parler français.

2.3.2. Échantillonnage

Pour déterminer le nombre nécessaire de participants, il est important de tenir compte du nombre de variables dans le modèle. Dans le cadre de ce projet, étant donné qu'il y a 6 variables (qualité de vie générale, qualité de vie au travail, âge, symptômes psychiatriques, estime de soi, satisfaction au travail) utilisées pour prédire le maintien en emploi, un échantillon de convenance de 72 participants a été ciblé. En effet, comme mentionné dans la première partie de ce chapitre, un échantillon de 60 participants est adéquat lorsqu'il y a 5 variables indépendantes incluses dans l'analyse de survie. Toutefois, en raison de la complexité du projet de recherche et de nombreuses difficultés de recrutement, un échantillon final de 67 participants ont pris part à l'étude.

2.3.3. Outils et collecte des données

Le déroulement de la collecte des données s'est fait en deux phases :

1. Phase 1 : une rencontre avec les participants afin de remplir les questionnaires (incluant la date d'origine).

2. Phase 2 : pour chaque participant, six mois après la rencontre en phase 1, un entretien téléphonique avec le directeur de l'entreprise sociale ou le responsable des ressources humaines a eu lieu afin de recueillir les informations sur le maintien en emploi du participant (incluant la date des dernières nouvelles). Les variables retenues dans le cadre de cette étude sont présentées dans le tableau 19.1.

Tableau 19.1.
Précision des variables retenues

Variable dépendante	Perte d'emploi (événement d'intérêt ciblé)	Dichotomique : 0) Non 1) Oui
	Durée du maintien en emploi (temps de survie)	Continue : nombre de jours
Variables indépendantes	Qualité de vie au travail (Lanctôt, Corbière et Durand, 2011)	Continue : (1 à 4)
	Qualité de vie en général (Lehman, 1995)	Continue : (1 à 7)
	Âge	Continue
	Symptômes psychiatriques (Derogatis et Melisaratos, 1983)	Continue : (1 à 5)
	Estime de soi en tant que travailleur (Corbière, Lanctôt, Sanquirgo et Lecomte, 2009)	Continue : (1 à 4)
	Satisfaction au travail (Weiss, Dawis, England et Lofquist, 1967)	Continue : (1 à 5)

2.4. Analyses des données et résultats

Des analyses de survie ont été conduites afin de déterminer les variables qui contribuent à expliquer le maintien en emploi des personnes avec un trouble mental grave travaillant au sein d'entreprises sociales québécoises. La version 17 du logiciel SPSS a été utilisée pour effectuer les analyses de cette recherche.

Comme mentionné précédemment, pour conduire une analyse de survie, les données doivent avoir un temps d'origine ainsi qu'une date de la survenue d'un événement. Dans le cas de cette étude, les données avaient comme date d'origine la date de la phase 1, et la date des dernières nouvelles (la date de l'événement d'intérêt, c'est-à-dire la date à laquelle le participant a perdu son emploi) ou la date de point (si le participant n'a pas perdu son emploi). Le modèle de Cox (1972) a été utilisé afin de modéliser la durée du maintien en emploi, entre le début de l'étude et la perte de l'emploi. Dans le cadre de ce modèle, la durée du maintien en emploi est considérée comme le temps de survie. Les participants qui seront toujours en emploi à la fin de l'étude seront considérés comme des observations censurées (une observation pour laquelle l'événement ne s'est pas produit). De plus, étant donné que le modèle de Cox permet de quantifier la contribution des diverses variables qui peuvent influencer le maintien en emploi, les covariables citées dans le tableau 19.1 ont été retenues dans le cadre de cette analyse.

Dans un premier temps, une estimation de la fonction de survie (courbe de survie) a été effectuée à l'aide de l'estimateur de Kaplan-Meier (1958). Cette méthode, qui est la plus utilisée, permet de calculer un taux de survie à chaque survenue d'un épisode de l'événement étudié. Ainsi, cette courbe de survie permet de déterminer la probabilité de se maintenir en emploi au-delà d'un certain temps, tout en sachant que la personne a réussi à se maintenir en emploi jusqu'à la fin de l'étude (date de point). Pour ce faire, les quatre étapes suivantes ont été effectuées dans le logiciel SPSS : 1) choisir *Analyse* → *Survie* → *Kaplan-Meier* ; 2) préciser l'*heure* (temps de survie, durée du maintien en emploi) et le *statut* (événement) ; 3) définir l'événement où 1 = perte d'emploi et 0 = maintien en emploi ; 4) sélectionner *Options* → cocher *Diagramme de survie*.

Les résultats de cette analyse sont premièrement illustrés dans un tableau récapitulatif (tableau 19.2). Celui-ci précise le nombre d'événements survenus (c.-à-d. le nombre de personnes qui ont perdu leur emploi). Ainsi, ce tableau indique que sept personnes ont perdu leur emploi au cours de l'étude. Dans le cadre de cette étude, aucun participant n'a été retiré ou ne s'est retiré de l'étude.

Tableau 19.2.
Récapitulatif du traitement des observations

n total	Nombre d'événements	Censuré	
		n	Pourcentage
67	7	60	89,6 %

Dans un deuxième temps, une table de survie est créée (tableau 19.3). Ce tableau indique pour tous les participants le temps de survie et si l'événement a eu lieu ou non. De façon plus précise, on constate que le premier participant a perdu son emploi après 21 jours et que le dernier participant l'a perdu après 171 jours. Ainsi, si l'on regarde l'estimation de la proportion de *survivants* cumulée, on constate que 89,6% des participants se sont maintenus en emploi après 171 jours.

Tableau 19.3.
Table de survie

	Nombre de jours	Maintien en emploi	Proportion de survivants cumulée à la période		Nombre d'événements cumulés	Nombre d'observations restantes
			Estimation	Erreur standard		
1	21	Non	0,985	0,015	1	66
2	62	Non	0,970	0,021	2	65
3	71	Non	0,955	0,025	3	64
4	81	Non	0,940	0,029	4	63
5	92	Non	0,925	0,032	5	62
6	115	Non	0,910	0,035	6	61
7	171	Non	0,896	0,037	7	60
8	181	Oui			7	59
9	181	Oui			7	58
10	181	Oui			7	57
…(67)	…	…	…	…	…	…

Par la suite, un tableau indique la moyenne et la médiane du délai de survie (tableau 19.4). Ainsi, les personnes se sont maintenues en emploi en moyenne 172 jours. Dans cet exemple, la médiane n'a pu être estimée toutefois, car moins de 50% de l'échantillon a vécu l'événement d'intérêt.

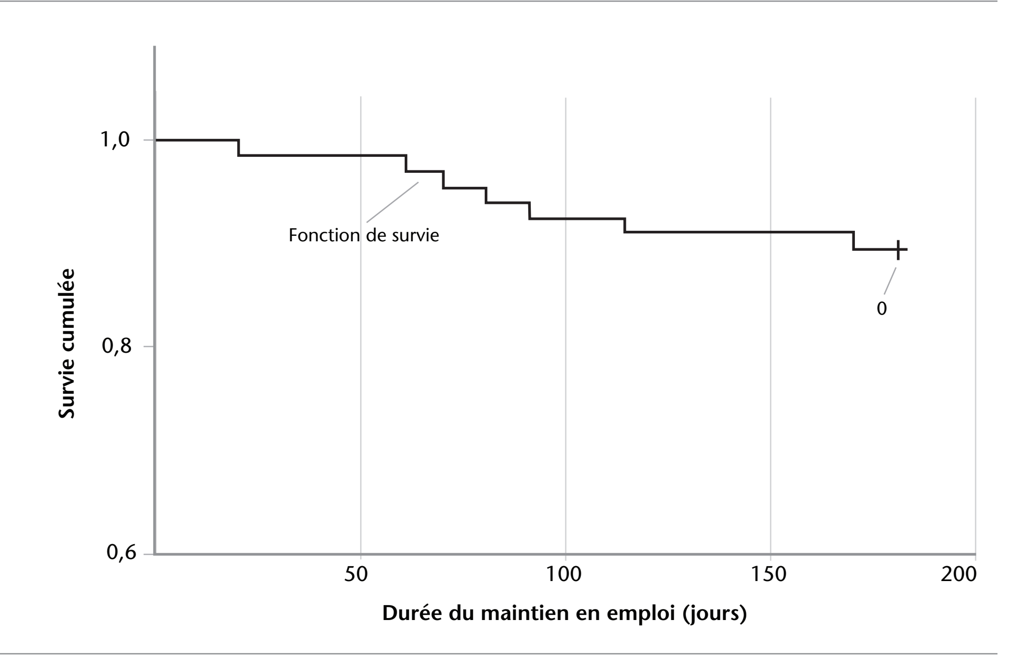
Tableau 19.4.
Moyenne et médiane du délai de survie

Moyenne[a]				Médiane			
Estimation	Erreur standard	Intervalle de confiance à 95 %		Estimation	Erreur standard	Intervalle de confiance à 95 %	
		Limite inférieure	Limite supérieure			Limite inférieure	Limite supérieure
172,134	3,921	164,448	179,820	.	.	.	.

a. L'estimation est limitée au délai de survie le plus long en cas de censure.

Enfin, la figure 19.6 illustre la fonction de survie. Cette figure est une façon différente de représenter les résultats de la table de survie. En effet, on y retrouve les mêmes informations, mais représentées de façon schématique. La première « marche d'escalier » arrive à 21, ce qui signifie que le premier participant a perdu son emploi après 21 jours. Étant donné que dans cette étude, sept personnes ont perdu leur emploi, on constate que l'escalier possède sept marches.

Figure 19.6.
Fonction de survie



Par la suite, une régression de Cox a été utilisée afin de modéliser la durée du maintien en emploi. Comme noté plus haut, ce modèle tient compte non seulement de la survenue ou non d'un événement, mais également du temps où l'événement survient. Dans le cadre de ce modèle, la durée du maintien en emploi a été considérée comme le temps de survie. Les participants qui étaient toujours en emploi lors de la phase 2 ont été considérés comme des observations censurées à droite (l'événement ne s'est pas encore produit à la fin de l'étude pour ces participants). De plus, le modèle de Cox a permis de quantifier la contribution des variables explicatives qui peuvent influencer le maintien en emploi. Pour ce faire, les six étapes suivantes ont été effectuées dans le logiciel SPSS : 1) choisir *Analyse* → *Survie* → *Modèle de Cox* ; 2) préciser l'*heure* (temps de survie), le *statut* (événement), et les *covariables* ; 3) définir l'événement où 1 = perte d'emploi et 0 = maintien en emploi ; 4) sélectionner *Nominales* pour déterminer les variables catégorielles et choisir la modalité de référence au choix (notez que si vous sélectionnez *Première*, vous devez cliquer sur *Changer* pour que le choix soit effectif) ; 5) sélectionner *Diagrammes* → cocher option *Survie* ; et finalement 6) sélectionner *Options* → cocher *CI pour exp(B)*.

Le tableau 19.5 indique toutes les variables qui ont été retenues dans l'équation. Ainsi, l'unique variable qui est significativement associée à la perte d'un emploi est la qualité de vie au travail ($p \leq 0{,}05$ et IC ne contient pas le 1). Plus précisément, les personnes qui avaient un plus haut niveau de qualité de vie au travail étaient plus susceptibles de maintenir leur emploi que celles qui avaient des niveaux inférieurs de qualité de vie au travail. Dit autrement, les personnes qui ont un haut niveau de qualité de vie au travail sont 99,5 % moins à risque de perdre leur emploi que celles qui ont un niveau de qualité de vie au travail plus faible. Pour arriver à ce résultat, on soustrait la valeur « Exp(β) de la variable qualité de vie au travail » à « 1 ». Ainsi, cela donne le pourcentage de valeur expliquée. Les résultats montrent que les autres variables ciblées ne contribuent pas de façon significative pour expliquer la perte d'emploi dans le contexte des entreprises sociales. Afin de présenter les résultats d'une analyse de régression de Cox, les informations suivantes sont généralement rapportées sous forme d'un tableau : nom des variables, valeur du coefficient β, le ratio des risques Exp(β) et son intervalle de confiance (tableau 19.5).

Tableau 19.5.
Variables retenues dans l'équation

	β	Erreur standard	Wald	ddl	Signif.	Exp (β)	95 % IC pour Exp(β)	
							Inférieure	Supérieure
Âge	–0,066	0,049	1,857	1	0,173	0,936	0,85	1,03
Qualité de vie au travail	–5,288	2,564	4,253	1	0,039	0,005	0,00	0,77
Estime de soi	2,102	1,266	2,756	1	0,097	8,18	0,68	97,86
Symptômes	306	0,678	0,204	1	0,652	1,36	0,36	5,13
Satisfaction au travail	2,121	1,500	1,999	1	0,157	8,34	0,44	157,72
Qualité de vie générale	–0,178	0,369	0,233	1	0,629	0,63	0,41	1,73

CONCLUSION

L'analyse de survie est une analyse relativement simple qui permet de tenir compte de l'aspect temporel dans le cadre d'une étude longitudinale. Une des méthodes les plus fréquemment utilisées est la courbe de survie de Kaplan-Meier (1958). De plus, par l'entremise de la régression de Cox (1972), il est possible de constater la contribution de variables contrôles explicatives, qualitatives ou quantitatives, dans un modèle prédictif. Cette analyse peut être utile pour l'étude de nombreux concepts reliés à la santé mentale. Par exemple, l'analyse de survie a été utilisée pour mieux comprendre la durée de l'absence de traitement psychotique (Compton *et al.*, 2008), pour étudier les durées d'absences au travail pour des raisons de maladies (Christensen *et al.*, 2007), pour évaluer les prédicteurs à long terme de la dépression (Godin *et al.*, 2009) et pour évaluer les répercussions de la maladie mentale sur le risque de perte d'emploi (Nelson et Kim, 2011). En résumé, la flexibilité de ce type d'analyse ouvre les portes à de nouvelles avenues de recherche intéressantes en santé mentale.

RÉFÉRENCES

Allison, P.D. (1995). *Survival Analysis Using the SAS System : A Practical Guide*, Cary, SAS Institute.

Bassett, J., C. Lloyd et H. Bassett (2001). « Work issues for young people with psychosis : Barries to employment », *British Journal of Occupational Therapy*, vol. 64, p. 66-72.

BÖHMER, P.E. (1912). *Theorie der unabhängigen Wahrscheinlichkeiten Rapports*, Communication présentée au Congrès international d'actuaires, Amsterdam.

BRESLOW, N. (1974). « Covariance analysis of censored survival data », *Biometrics*, vol. 30, p. 88-99.

CATTY, J., P. LISSOUBA, S. WHITE, T. BECKER, R.E. DRAKE et A. FIORITTI (2008). « Predictors of employment for people with severe mental illness : Results of an international six-centre randomised controlled trial », *British Journal of Psychiatry*, vol. 192, nº 3, p. 224-231.

CHRISTENSEN, K.B., P. ANDERSEN, L. SMITH-HANSEN, M.L. NIELSEN et T.S. KRISTENSEN (2007). « Analyzing sickness absence with statistical models for survival data », *Scandinavian Journal of Work, Environment and Health*, vol. 33, nº 3, p. 233-239.

COMPTON, M.T., V.H. CHIEN, A.S. LEINER, S.M. GOULDING et P.S. WEISS (2008). « Mode of onset of psychosis and family involvement in help-seeking as determinants of duration of untreated psychosis », *Social Psychiatry and Psychiatric Epidemiology*, vol. 43, p. 975-982.

CORBIÈRE, M., N. LANCTÔT, N. SANQUIRGO et T. LECOMTE (2009). « Evaluation of self-esteem as a worker for people with severe mental disorders », *Journal of Vocational Rehabilitation*, vol. 30, p. 87-98.

CORBIÈRE, M., T. LECOMTE et N. LANCTÔT (2011). « Services de réintégration au travail et interventions ponctuelles en réadaptation pour les personnes avec un trouble mental grave », dans M. Corbière et M.-J. Durand (dir.), *Du trouble mental à l'incapacité au travail*, Québec, Presses de l'Université du Québec, p. 225-252.

CORBIÈRE, M., A. LESAGE, K. VILLENEUVE et C. MERCIER (2006). « Le maintien en emploi de personnes souffrant d'une maladie mentale », *Santé mentale au Québec*, vol. 31, nº 2, p. 215-235.

COX, D.R. (1972). « Regression models and life tables (with discussion) », *Journal of the Royal Statistical Society, Series B*, vol. 34, p. 187-220.

DELONG, D.M., G.H. GUIRGUIS et Y.C. SO (1994). « Efficient computation of subset selection probabilities with application to Cox regression », *Biometrika*, vol. 81, p. 607-611.

DEROGATIS, L.R. et N. MELISARATOS (1983). « The Brief Symptom Inventory : An introduction report », *Psychological Medicine*, vol. 13, nº 3, p. 595-605.

DUMAIS, L. et Y. VAILLANCOURT (2005). « L'économie sociale et l'inclusion des personnes vulnérables au Québec », *Le Partenaire*, vol. 12, nº 2, p. 4-10.

EFRON, B. (1977). « The efficiency of Cox's likelihood function for censored data », *Journal of the American Statistical Association*, vol. 72, p. 557-565.

ELIASON, S.R. (1993). *Maximum Likelihood Estimation : Logic and Practice*, Newbury Park, Sage Publications.

GODIN, I., M. KORNITZER, N. CLUMECK, P. LINKOWSKI, F. VALENTE et F. KITTEL (2009). « Gender specificity in the prediction of clinically diagnosed depression : Results of a large cohort of Belgian workers », *Social Psychiatry and Psychiatric Epidemiology*, vol. 44, p. 592-600.

KALBLFLEISCH, J.D. et R.L. PRENTICE (1980). *The Statistical Analysis of Failure Time Data*, New York, Wiley.

KAPLAN, E.L. et P. MEIER (1958). « Nonparametric estimation from incomplete observations », *Journal of the American Statistical Association*, vol. 53, p. 457-481.

LANCTÔT, N., M. CORBIÈRE et M.-J. DURAND (2011). *Questionnaire sur la qualité de vie au travail*, Sherbrooke, Université de Sherbrooke.

LANCTÔT, N., M. CORBIÈRE et M.-J. DURAND (2012). « Job tenure and quality of work life of people with psychiatric disabilities working in social enterprises », *Journal of Vocational Rehabilitation*, vol. 37, nº 1, p. 39-48.

LEHMAN, A.F. (1995). *Evaluating Quality of Life for Persons with Severe Mental Illness : Assessment Toolkit*, Cambridge, Evaluation Center at Health Services Research Institute.

MUESER, K.T., M.P. SALYERS et P.R. MUESER (2001). « A prospective analysis of work in schizophrenia », *Schizophrenia Bulletin*, vol. 27, p. 281-296.

NELSON, R.E. et J. KIM (2011). « The impact of mental illness on the risk of employment termination », *Journal of Mental Health Policy and Economics*, vol. 14, nº 1, p. 39-52.

RESNICK, S.G. et G.R. BOND (2001). « The Indiana Job Satisfaction Scale : Job satisfaction in vocational rehabilitation for people with severe mental illness », *Psychiatric Rehabilitation Journal*, vol. 25, nº 1, p. 12-19.

WEISS, D.J., R.V. DAWIS, G.W. ENGLAND et L.H. LOFQUIST (1967). *Manual for the Minnesota Satisfaction Questionnaire*, Minneapolis, University of Minnesota Industrial Relations Center.

WERWIORSKI, N.J. et E. FABIAN (2004). « Association between demographic and diagnostic factors and employment outcomes for people with psychiatric disabilities : A synthesis of recent research », *Mental Health Services Research*, vol. 6, nº 1, p. 9-21.

XIE, H., B.J. DAIN, D.R. BECKER et R.E. DRAKE (1997). « Job tenure among persons with severe mental illness », *Rehabilitation Counseling Bulletin*, vol. 40, nº 4, p. 230-240.



LES ANALYSES MULTINIVEAUX
Illustration à l'aide d'une étude évaluant l'efficacité d'une intervention scolaire ciblant les problèmes d'attention en classe et les difficultés de lecture

Marie-Christine Brault
Éric Dion
Véronique Dupéré

FORCES

- Elles corrigent le problème de dépendance des observations associé aux données ayant une structure nichée, c'est-à-dire une structure emboîtée, où les unités plus petites (personnes) sont regroupées au sein d'unités de niveau supérieur (groupes, organisations).
- Elles permettent la modélisation simultanée des relations aux niveaux micro et macro (p. ex., la personne et l'organisation), en plus des interactions entre ces niveaux.
- Elles s'utilisent autant avec les données longitudinales et transversales qu'avec les distributions normales et non normales.

LIMITES

- Elles sont d'un niveau de complexité élevé.
- Les données doivent posséder une structure hiérarchique, au sujet de laquelle l'information doit être disponible.
- Elles requièrent un nombre suffisant de personnes dans chaque groupe, ainsi qu'un nombre suffisant de groupes.

Dans les études portant sur la santé mentale, les participants sont souvent recrutés au sein d'organisations, comme les cliniques médicales, les milieux de travail ou les écoles. L'inclusion de participants provenant d'une même organisation représente à la fois une opportunité de recherche et un défi pour l'analyse. Sur le plan des opportunités, l'inclusion de plusieurs milieux et de plusieurs participants au sein de chacun de ces milieux permet d'étudier les liens entre les caractéristiques de ces environnements, comme le climat de travail, et la santé mentale des participants. Sur le plan des défis, l'analyse de données nichées, où des participants appartiennent à des organisations communes, requiert l'utilisation de techniques d'analyses particulières, comme l'analyse multiniveaux, puisque les observations dans ce contexte ne peuvent pas être considérées comme indépendantes. Ceci représente un problème, puisque l'indépendance des observations est l'un des postulats de base des analyses statistiques plus classiques, telle la régression linéaire. L'analyse multiniveaux est taillée sur mesure pour exploiter les forces des échantillons « multiniveaux » où des personnes sont nichées dans des organisations, tout en éliminant les problèmes liés au postulat d'indépendance des données.

Ce chapitre propose une présentation sommaire et appliquée du modèle linéaire hiérarchique relevant de l'analyse multiniveaux. D'abord, les éléments théoriques centraux et les particularités de cette analyse seront exposées, puis la procédure et l'interprétation des résultats seront illustrées à l'aide d'un exemple concret.

1. ÉLÉMENTS THÉORIQUES CENTRAUX DE L'ANALYSE MULTINIVEAUX

1.1. Structures nichées et postulat d'indépendance

1.1.1. Structures nichées

Les données possédant une structure nichée sont très courantes dans le domaine de la santé mentale. Le caractère niché peut provenir du fait que plusieurs participants à une étude font partie d'un même contexte ou d'une même organisation, ou encore du fait que plusieurs mesures sont recueillies auprès d'une même personne au fil du temps, dans le cadre d'études longitudinales (mesures répétées nichées dans une personne) (Diez-Roux, 2002). Dans ces deux cas, les observations sont dépendantes les unes des autres, puisqu'il est attendu que les mesures obtenues auprès d'une même personne au fil du temps sont corrélées entre elles (Dupéré *et al.*, 2007)

et que les observations provenant de personnes faisant partie d'un même groupe peuvent se ressembler davantage que celles provenant de groupes différents (Hox et Kreft, 1994).

Les études portant sur l'étiologie et l'épidémiologie sociales des troubles mentaux fournissent de bons exemples de structures nichées, car elles s'intéressent entre autres aux répercussions des caractéristiques de l'environnement social sur la santé des personnes. Dans ce contexte, il arrive fréquemment que les personnes soient regroupées selon leur quartier de résidence (Diez-Roux, 2000), selon les groupes ou organisations sociales auxquels elles appartiennent (famille, école, travail, etc.) (Allardyce et Boydell, 2006; Kidger *et al.*, 2012; Brière *et al.*, 2013), voire selon leur cohorte de naissance (Yang, 2007; Brault *et al.*, 2012). Le point commun de ces études est la volonté de dissocier les effets des facteurs de risque ou de protection contextuels de ceux relevant de la personne. Ces études s'intéressent aussi typiquement aux interactions entre ces facteurs de risque aux niveaux contextuel et individuel.

Les études cherchant à établir l'efficacité d'un traitement de groupe à l'aide d'un devis expérimental de nature longitudinale ont également une structure nichée. D'abord, les mesures répétées sont nichées dans les personnes, ensuite, ces personnes sont nichées dans les groupes d'intervention. L'étude de Lecomte et collaborateurs (2008), en psychiatrie, en constitue un bon exemple. Ces chercheurs ont suivi 130 adultes présentant des symptômes psychotiques persistants sur une période de 15 mois avec pour objectif de comparer l'efficacité d'une thérapie cognitive-comportementale à celle d'une intervention plus traditionnelle portant sur les habiletés sociales (voir le chapitre 10 de cet ouvrage pour plus de détails sur cette étude). En plus de devoir considérer la nature longitudinale du suivi thérapeutique (trois temps de mesure), cette étude a dû prendre en compte les conditions expérimentales (caractéristiques des thérapeutes, de l'environnement et du moment de l'intervention). À cette fin, les auteurs ont utilisé un modèle à trois niveaux, où les mesures répétées (niveau 1) sont nichées dans les personnes (niveau 2), qui à leur tour sont nichées dans les groupes d'intervention (niveau 3).

1.1.2. Postulat d'indépendance

Nous avons vu ci-dessus qu'il existe de nombreuses situations où les chercheurs du domaine de la santé mentale souhaitent (ou doivent) tenir compte de l'environnement – école, organisation sociale, quartier de résidence, hôpital, médecins, groupe d'intervention, etc. – dans lequel évoluent (sont nichés) les personnes ou les phénomènes observés. Dans tous ces

cas, il est fort probable que le postulat d'indépendance des observations ne soit pas respecté, ce qui affectera négativement la validité des résultats. Selon le postulat d'indépendance, il est possible de tirer des conclusions au sujet d'une population uniquement si les réponses obtenues auprès de chaque participant composant l'échantillon n'influencent pas les réponses des autres participants au sein de cet échantillon. Par exemple, si l'on veut estimer la prévalence du diagnostic de trouble du déficit de l'attention/hyperactivité (TDAH) chez les enfants québécois, il peut être problématique d'inclure des enfants issus d'une même fratrie dans l'étude, puisque le TDAH apparaît comme l'un des troubles psychiatriques les plus héréditaires (Faraone *et al.*, 2005). En d'autres termes, les informations obtenues auprès d'enfants partageant un lien de parenté ne peuvent pas être considérées comme indépendantes puisque la probabilité de recevoir ce diagnostic augmente si un membre de la famille l'a déjà obtenu.

À cet effet, la dépendance des observations a pour conséquence d'augmenter la probabilité de rapporter comme significatives des différences ou des relations qui n'existent pas dans les faits (Hox et Kreft, 1994 ; Bickel, 2007). Les analyses traditionnelles ne tenant pas compte de la structure hiérarchique des observations augmenteront également le risque d'attribuer au niveau agrégé des effets observés au niveau individuel (erreur atomistique) ou au contraire d'attribuer au niveau individuel des effets observés au niveau agrégé (erreur écologique) (Diez-Roux, 1998 ; Snijders et Bosker, 1999). Notamment pour ces raisons, mais aussi parce que les analyses multiniveaux sont maintenant bien connues des évaluateurs, il est devenu difficile de faire accepter un article dans lequel des analyses conventionnelles ont été appliquées à des données ayant de toute évidence une structure hiérarchique.

1.1.3. Justification du recours à l'analyse multiniveaux

En raison de la complexité statistique de l'analyse multiniveaux (Diez-Roux, 1998), il est indispensable de s'assurer de sa nécessité avant de l'utiliser. À cet effet, le choix de recourir à l'analyse multiniveaux doit être justifié par la présence de deux conditions essentielles. D'abord, il est nécessaire d'avoir des données nichées (Bickel, 2007), peu importe qu'il s'agisse de personnes nichées dans des contextes ou de mesures répétées nichées dans des personnes. Ensuite, il doit y avoir une dépendance réelle des observations caractérisée par une homogénéité intragroupe (Bickel, 2007 ; Peugh, 2010). La vérification de cette condition se fait grâce au calcul du coefficient de corrélation intraclasse, qui reflète le degré de ressemblance des observations d'un même groupe (Diez-Roux, 2002). La procédure à suivre sera illustrée en détail dans l'exemple (section 2.3).

1.2. Particularités et procédures de l'analyse multiniveaux

1.2.1. Objectifs de l'analyse multiniveaux

L'analyse multiniveaux permet de modéliser plusieurs types de relation micro/macro, dont les interactions entre les niveaux (Snijders et Bosker, 1999). Elle est utilisée en particulier pour partager la variance entre les différents niveaux hiérarchiques, sans nécessiter l'agrégation des données ou la séparation des analyses par niveau (Raudenbush et Bryk, 2002). Concrètement, ceci permet d'évaluer l'influence relative des différents niveaux sur la variable dépendante en utilisant toute l'information disponible et sans perdre de puissance statistique (l'agrégation des données diminue la puissance en diminuant le nombre de cas). Elle permet aussi de déterminer si les différences s'observent surtout entre les groupes (inter) ou au sein des groupes (intra).

1.2.2. Coefficients fixes et aléatoires

Synonyme de modèles hiérarchiques linéaires (Raudenbush et Bryk, 2002), de modèles de régression hiérarchique (Bickel, 2007) ou de modèles à coefficients aléatoires (voir Diez-Roux, 2002, pour un résumé explicatif des différentes terminologies), l'analyse multiniveaux peut être conçue comme une version plus complexe de l'analyse de régression conventionnelle (Bickel, 2007). En fait, il s'agit d'un type de modèle mixte, car il inclut à la fois des coefficients fixes – qui supposent que l'effet sera le même peu importe le contexte (p. ex., être de sexe féminin a le même effet sur la réussite scolaire peu importe l'école) – et des coefficients aléatoires – qui anticipent un effet propre à chacun des contextes, par exemple, une relation différente entre le statut socioéconomique de l'élève et sa réussite scolaire dans chaque école (Raudenbush et Bryk, 2002). La présence de coefficients aléatoires dans l'analyse multiniveaux permet d'obtenir une estimation adéquate ou juste des coefficients de régression, qui tient compte de la structure hiérarchique des données (Raudenbush et Bryk, 2002) et par conséquent de leur corrélation (Diez-Roux, 2002 ; Dupéré *et al.*, 2007).

1.2.3. Types d'analyses multiniveaux

Il existe une variété de modèles multiniveaux, selon les caractéristiques des données et de leur structure. Les données longitudinales requièrent un modèle que l'on nomme courbe de croissance (Boyle et Willms, 2001 ; Singer et Willett, 2003), alors que certaines structures nichées peuvent nécessiter des modèles à trois (et plus) niveaux (Snijders et Bosker, 1999 ;

Raudenbush et Bryk, 2002) ou des modèles à double classification, où les personnes relèvent de deux niveaux supérieurs (p. ex., élèves nichés à la fois dans des écoles et dans des quartiers) (Raudenbush, 1993; Yang et Land, 2006). Ces modèles plus complexes ne pourront être abordés ici; l'exemple présenté dans la prochaine section de ce chapitre mettra plutôt l'accent sur les modèles multiniveaux transversaux, où les personnes sont nichées dans un contexte. L'exemple typique d'une structure hiérarchique provient du domaine de l'éducation, où les élèves sont nichés dans des classes, qui sont à leur tour nichées dans une école, qui elle fait partie d'une commission scolaire, etc. (Raudenbush et Bryk, 2002). Cette série de regroupements représente autant de niveaux, du plus spécifique (l'élève) au plus général (la commission scolaire). Par exemple, l'analyse multiniveaux permet de départager à quel point les différences sur le plan des symptômes de TDAH sont dues à des différences individuelles (niveau 1) ou à l'école fréquentée (niveau 2).

1.2.4. Équations

Comme pour les modèles de régressions conventionnelles, la représentation des équations qui serviront de base à chacun des modèles multiniveaux est une étape cruciale qui permet de s'assurer que la modélisation statistique est complète et exempte d'erreur. Une des particularités des équations des modèles multiniveaux est l'emploi des paramètres estimés au premier niveau (p. ex., β_{0j} et β_{1j}) comme variables dépendantes au deuxième niveau d'analyse (Raudenbush et Bryk, 2002). De plus, chacune des équations est composée d'une partie présentant les effets fixes et d'une autre présentant les effets aléatoires. Il existe plusieurs manières de rédiger les équations (voir Singer 1998), mais d'un point de vue pédagogique, il nous semble plus adéquat de mettre l'accent sur les équations séparées pour chacun des niveaux. Les équations présentées dans l'illustration de ce chapitre utilisent la nomenclature de Raudenbush et Bryk (2002).

1.2.5. Taille requise des échantillons

La puissance statistique en multiniveaux dépend, entre autres, de la taille de chacun des niveaux (Snijders, 2005). Les connaissances au sujet du nombre minimal d'unités nécessaires par niveau sont toutefois limitées (Bickel, 2007). Il est généralement admis que la plus grande restriction provient de la taille de l'échantillon au niveau supérieur (Maas et Hox, 2005; Snijders, 2005). À cet effet, bien qu'une taille de 10 (Snijders et Bosker, 1999) à 30 groupes au niveau supérieur soit considérée comme correcte par certains (voir Bickel, 2007, pour une discussion), une taille de plus de 50 cas au

niveau supérieur est préférable pour minimiser les biais (Maas et Hox, 2005). Cela correspondrait par exemple à 50 écoles, dans une étude où des élèves sont nichés dans des écoles. Au niveau inférieur, un ratio de 30 observations (élèves) par groupe (école) est habituellement recommandé (Bickel, 2007), bien qu'un plus petit ratio (5 observations par groupe) soit également acceptable (Maas et Hox, 2005). Au final, la détermination de la taille de l'échantillon devrait toujours tenir compte des objectifs des analyses et du choix des paramètres d'intérêt (Snijders, 2005).

1.2.6. Procédures

Afin de permettre l'interprétation correcte des paramètres de l'analyse multiniveaux, il est nécessaire que les variables, principalement celles relevant du niveau 1, soient au préalable « centrées ». Le centrage des variables permet d'attribuer une signification claire et concrète au score zéro, qui est trop souvent mal défini ou tout simplement inexistant (Enders et Tofighi, 2007). Il est important de choisir soigneusement le type de centrage, car celui-ci a des répercussions sur l'interprétation effectuée (Enders et Tofighi, 2007). Il s'effectue en choisissant l'une des deux possibilités suivantes : *a*) le centrage sur la grande moyenne, où un score de zéro correspond désormais à la moyenne de l'échantillon complet, ou *b*) le centrage sur la moyenne du groupe, où la signification du zéro varie d'un groupe à l'autre. Cette dernière option de centrage complexifie l'interprétation des résultats et elle n'est habituellement pas retenue.

L'analyse multiniveaux débute habituellement par l'estimation d'un modèle inconditionnel qui n'inclut aucune variable dépendante. Ce modèle permet de décomposer la variance de la variable dépendante entre les divers niveaux d'analyse et d'estimer le coefficient de corrélation intraclasse, indiquant le degré de ressemblance des personnes dans un même groupe (Singer, 1998 ; Raudenbush et Bryk, 2002). Puis, les variables indépendantes sont introduites graduellement, par bloc, dans un ordre qui est fonction des hypothèses. Cette introduction graduelle permet de déterminer comment se comporte le modèle, mais aussi d'examiner l'effet unique de chacun des blocs. En pratique, l'insertion de ces variables se fait en suivant une procédure similaire à celle habituellement utilisée dans le cadre de régressions conventionnelles, mais en utilisant un programme permettant de spécifier le niveau de chacune des variables insérées.

1.2.7. Logiciels

Le choix de logiciels permettant l'analyse multiniveaux est varié. Alors que certains (HLM, MLwiN) ont été spécialement conçus pour cette analyse, la plupart des logiciels statistiques connus (SPSS, SAS, Stata, Mplus, etc.) intègrent une procédure qui permet l'analyse multiniveaux. Les références suivantes informeront le lecteur sur l'analyse multiniveaux effectuée à l'aide de SAS (Singer, 1998), de SPSS (Bickel, 2007) et de HLM (Raudenbush *et al.*, 2004).

1.2.8. Données manquantes

Malgré les avancées en matière de gestion des données manquantes, les logiciels permettant l'analyse multiniveaux continuent par défaut de retirer de l'analyse les sujets ayant des observations manquantes au niveau 1 et ceux faisant partie d'un groupe (niveau supérieur) pour lequel des observations sont manquantes. Puisque cette procédure peut créer des biais (van Buuren, 2010), il est fortement recommandé de choisir un autre mode de gestion des données manquantes. Cette discussion dépasse toutefois la portée du présent chapitre. Le lecteur intéressé par la gestion des données manquantes pour l'analyse multiniveaux pourra consulter van Buuren (2010).

2. ILLUSTRATION D'UNE ANALYSE MULTINIVEAUX

Nous reprenons ici les données de l'étude de Dion et collaborateurs (2011) pour présenter un exemple d'analyse multiniveaux classique utilisant le modèle linéaire hiérarchique, où les élèves sont nichés dans des écoles. Dans cette étude, les chercheurs ont testé, à l'aide d'un devis expérimental, l'efficacité d'une intervention visant à améliorer la capacité d'attention en classe et à prévenir les problèmes de lecture des élèves de première année.

2.1. Objectif de recherche

Les problèmes d'attention en classe, qu'ils soient ou non liés au diagnostic de TDAH, figurent parmi les plus importants facteurs de risque des difficultés en lecture (p. ex., Dally, 2006). Par ailleurs, les élèves inattentifs ne semblent pas bénéficier autant que leurs pairs des interventions éducatives ciblant les habiletés de lecture (Al Otaiba et Fuchs, 2002 ; Rabiner et Malone, 2004). Les raisons de s'intéresser à l'attention en classe vont bien au-delà de ses conséquences sur les habiletés en lecture. En effet, des études ont

démontré que les comportements d'inattention dans l'enfance étaient associés notamment à un risque accru de décrochage scolaire (Pingault *et al.*, 2011) et de dépendance à la nicotine à l'âge adulte (Pingault *et al.*, 2012).

Bien que la prescription de psychostimulants soit en hausse chez les enfants d'âge scolaire pour traiter les symptômes d'inattention et d'hyperactivité associés au diagnostic de TDAH (Brault et Lacourse, 2012), plusieurs effets secondaires ont été rapportés (Hosenbocus et Chahal, 2009). Il est envisageable qu'une intervention implantée par l'enseignant puisse améliorer l'attention en classe de plusieurs élèves. À cet effet, les données illustratives présentées ici permettent de comparer l'efficacité d'interventions ciblant *a*) les habiletés en lecture exclusivement ou *b*) à la fois ces habiletés et l'attention en classe.

2.2. Description de l'échantillon, du devis expérimental et des variables

2.2.1. Échantillon

Dans chacune des 58 classes participant au projet, 7 enfants ont été sélectionnés sur la base de leurs habiletés émergentes en lecture et de leur niveau d'attention. Des élèves faibles et moyens en lecture ont été sélectionnés en s'assurant d'inclure tous les élèves inattentifs. L'analyse de cas complets a été privilégiée pour les besoins de cette démonstration, alors après exclusion des données manquantes, l'échantillon utilisé pour cet exemple compte 375 élèves nichés dans 30 écoles[1] primaires francophones de la région de Montréal, des écoles principalement situées dans des quartiers à très faible niveau socioéconomique.

2.2.2. Devis expérimental et interventions

Les écoles ont été aléatoirement assignées à l'une des trois conditions expérimentales suivantes : témoin, intervention unique et intervention combinée. Les enseignants des écoles du groupe témoin ont continué à enseigner la lecture et à gérer leur groupe comme d'habitude. Les enseignants des écoles de l'intervention unique continuaient à gérer leur classe comme d'habitude, mais devaient enseigner l'apprentissage de la lecture à l'aide d'activités de tutorat entre pairs *Apprendre à lire à deux* (Dion *et al.*, 2005a). Les enseignants de l'intervention combinée devaient utiliser ces mêmes activités de

1. Ce nombre d'unités de niveau supérieur est en deçà de celui suggéré précédemment. La puissance statistique et la taille de l'effet associées à ce petit *n* sont discutées dans l'article original (Dion *et al.*, 2011).

tutorat entre pairs pour l'apprentissage de la lecture, en plus d'implanter l'intervention *Attention, je lis !* (Dion *et al.*, 2005b), qui permet d'améliorer l'attention en classe.

2.2.3. Variables

Les informations de base au sujet des variables figurent dans le tableau 20.1. La variable dépendante est la proportion de comportements d'attention optimale observée en classe (ATT_OP) telle que mesurée durant une période consacrée à l'enseignement de la lecture en classe. La variable CON_ATTC correspond au score d'attention à l'échelle de Conners (2000) remplie par l'enseignant. Afin de respecter les recommandations de l'analyse multi-niveaux et ainsi de faciliter l'interprétation (Enders et Tofighi, 2007), cette variable a été centrée sur la grande moyenne : un score de zéro correspond désormais à la moyenne (11,35) pour l'ensemble de l'échantillon. La variable *Condition* est une variable nominale à trois catégories qui représentent les

Tableau 20.1.
Description des variables tirées de l'étude expérimentale de Dion *et al*. (2011)

Nom de la variable	Description	Mesure/Codification
Niveau 1 – élève		
ATT_OP	Proportion de comportements d'attention optimale observée en classe	Variable continue (moyenne = 0,725 ; écart-type = 0,232)
CON_ATTC	Score d'attention au prétest sur l'échelle de Connors (2002)	Variable continue, centrée sur la grande moyenne (moyenne = 0 ; écart-type = 4,34) Score élevé = problème d'attention
Niveau 2 – école		
Condition	Condition expérimentale	0 : Témoin 1 : Intervention unique (*Apprendre à lire à deux*) 2 : Intervention combinée (*Apprendre à lire à deux + Attention je lis !*)
C_INTERV	Variable factice binaire représentant la condition expérimentale	0 : Témoin (*Condition* = 0) 1 : Intervention (*Condition* = 1 ou 2)
C_AJL	Variable factice binaire représentant la condition expérimentale	0 : Témoin + Intervention unique (*Condition* = 0 ou 1) 1 : Intervention combinée (*Condition* = 2)

conditions expérimentales. Comme pour l'analyse de régression traditionnelle, les variables nominales, pour être introduites dans l'analyse, doivent être préalablement transformées en plusieurs variables factices binaires (Fox, 1999). Pour les besoins de l'analyse, nous avons créé deux variables (une de moins que le nombre de catégories de la variable originale) : *C_INTERV* et *C_AJL*. Comme expliqué dans le tableau 20.1, la variable *C_INTERV* établit la distinction entre les écoles de la condition témoin et les autres (celles qui reçoivent une forme d'intervention ou une autre), alors que la variable *C_AJL* distingue les autres écoles où l'intervention visant à améliorer l'attention a été implantée de celles où elle ne l'a pas été. Pour une description plus détaillée des variables, le lecteur est invité à consulter l'article original (Dion *et al.*, 2011).

2.3. Analyse multiniveaux : procédure et équations

L'analyse est réalisée à l'aide d'un modèle à deux niveaux dans le cadre duquel les élèves (niveau 1) sont nichés dans les écoles (niveau 2), et où la condition expérimentale est considérée comme une caractéristique de l'école puisque c'est cette dernière qui a été assignée aléatoirement aux conditions (tous les enseignants participants de l'école étaient assignés à la même condition). Comme suggéré précédemment, nous avons débuté par l'estimation d'un modèle inconditionnel qui n'inclut aucune variable indépendante (voir équations 1 et 2).

Avec $i = 1, 2, \ldots n_j$ élèves à l'intérieur d'une école j et $j = 1, 2, \ldots$ J écoles, l'équation de niveau 1 (élève) et l'équation de niveau 2 (écoles) sont les suivantes :

$$Y_{ij} = \beta_{0j} + r_{ij} \quad (1)$$

$$\beta_{0j} = \gamma_{00} + \mu_{0j} \quad (2)$$

où

- Y_{ij} = la proportion de comportements d'attention optimale observée en classe pour l'élève i de l'école j ;
- β_{0j} = la proportion moyenne de comportements d'attention optimale observée en classe (ordonnée à l'origine) pour l'école j ;
- γ_{00} = la proportion moyenne de comportements d'attention optimale observée en classe (ordonnée à l'origine) pour les écoles de l'échantillon ;
- r_{ij} et μ_{0j} = les paramètres aléatoires de niveau 1 et 2. Tous deux sont distribués normalement avec une moyenne de zéro et une variance σ^2 pour le niveau 1 et τ_{00} pour le niveau 2.

Ce modèle inconditionnel permet de calculer le coefficient de corrélation intraclasse, dont la formule est la suivante :

$$\tau_{00}/\sigma^2 + \tau_{00} \quad (3)$$

Comme présenté à l'équation 3, le coefficient de corrélation intraclasse s'obtient en divisant la variance intergroupe (τ_{00}, représentée par la ligne Intercept dans le tableau 20.2) par la variance totale ($\sigma^2 + \tau_{00}$) composée de la variabilité intragroupe et intergroupe (σ^2, représentée par la ligne Residual dans le tableau 20.2). Plus les groupes sont homogènes et distincts les uns des autres, plus la valeur du coefficient est élevée. En effectuant le calcul à partir des résultats de notre exemple – 0,017/(0,037 + 0,017) – nous obtenons un coefficient de corrélation intraclasse[2] de 0,315, indiquant que 31,5 % de la variance de la variable dépendante se situe entre les écoles. L'ajout de variables indépendantes au second niveau d'analyse (p. ex., nos variables de conditions expérimentales) permettra d'expliquer cette variabilité intergroupe, au moins en partie.

Tableau 20.2.
Résultats du modèle nul – partie aléatoire

Valeurs estimées des paramètres de covariance					
Param de cov	**Sujet**	**Valeur estimée**	**Erreur type**	**Valeur Z**	**Pr > Z**
Intercept	école	0,017	0,005	3,25	0,001
Residual		0,037	0,003	13,14	< 0,000

Cette étape de la modélisation nous a permis de justifier l'utilisation de l'analyse multiniveaux. En effet, comme l'indique la valeur élevée du coefficient de corrélation intraclasse, les élèves regroupés dans une même école se ressemblent davantage que les élèves des autres écoles. À noter que la corrélation intraclasse doit vraiment être faible pour que des analyses conventionnelles puissent être utilisées (Bickel, 2007).

2.3.1. Modèle final

Nous avons ensuite introduit les variables indépendantes graduellement, par bloc, dans un ordre qui est fonction des hypothèses. Dans le cas qui nous intéresse, nous avons d'abord entré les variables de niveau école, puis les variables individuelles. Pour des raisons de concision, seules les

2. Le coefficient de corrélation intraclasse indique la proportion de la variance de la variable dépendante qui se situe entre les unités du niveau supérieur. Il est habituellement rapporté en pourcentage.

équations associées au modèle final sont présentées ici. Avec $i = 1,2, \ldots n_j$ élèves à l'intérieur d'une école j et $j = 1, 2, \ldots$ J écoles, l'équation de niveau 1 (élève) est la suivante :

$$Y_{ij} = \beta_{0j} + \beta_{1j}\ (\text{CON_ATTC})_{ij} + r_{ij} \quad (4)$$

où

Y_{ij} = la proportion de comportements d'attention optimale observée en classe pour l'élève i de l'école j ;

β_{0j} = la proportion moyenne de comportements d'attention optimale observée en classe (ordonnée à l'origine) pour l'école j ;

β_{1j} = la relation entre la proportion de comportements d'attention optimale observée en classe et le score d'attention au prétest (*CON_ATTC*) ;

et r_{ij} = le paramètre aléatoire qui représente l'écart entre l'attention optimale prédite et observée pour chaque élève.

Les équations pour le second niveau (école) sont :

$$B_{0j} = \gamma_{00} + \gamma_{01}(\text{C_INTERV})_j + \gamma_{02}(\text{C_AJL})_j + \mu_{0j} \quad (5)$$

où

$$B_{1j} = \gamma_{10}$$

γ_{00} = la proportion moyenne de comportements d'attention optimale observée en classe (ordonnée à l'origine) pour les écoles du groupe témoin ;

γ_{01} = la différence de score de la proportion moyenne de comportements d'attention optimale observée en classe entre les écoles du groupe témoin et celles recevant l'intervention en lecture (*C_INTERV*) ;

γ_{02} = la différence entre les écoles recevant simplement l'intervention en lecture et celles recevant en plus l'intervention visant à améliorer l'attention (*C_AJL*) ;

μ_{0j} = le paramètre aléatoire de niveau 2 qui représente l'écart entre l'attention optimale prédite et observée pour chaque école ; et

γ_{10} = le changement moyen (la pente) de la proportion moyenne de comportements d'attention optimale observée en classe qui est associé au niveau d'attention au prétest.

Il est important de noter que nous avons volontairement omis un terme d'erreur (μ_{1j}) dans l'équation du β_{1j}, ce qui fait en sorte que la relation entre le score d'attention au prétest (*CON_ATTC*) et la proportion

moyenne de comportements d'attention optimale observée en classe ne varie pas d'une école à l'autre (la relation est considérée comme fixe plutôt que comme aléatoire).

2.4. Analyse multiniveaux : interprétations des tableaux de résultats

Les résultats ont été obtenus à l'aide de la procédure PROC MIXED du logiciel SAS, version 9.3 (SAS Institute Inc., 2002-2010). Les tableaux de résultats qui sont les plus importants concernent ceux des effets fixes et des effets aléatoires.

Les résultats de la partie aléatoire du modèle final sont présentés au tableau 20.3. La variance intergroupe (représentée par la ligne Intercept : 0,011, $p \leq 0{,}05$) est significative. Cela indique que les écoles se distinguent les unes des autres sur leur score initial moyen de proportion de comportements d'attention optimale observée en classe. La proportion de comportements d'attention optimale observée en classe relève toutefois davantage du niveau de l'élève (représentée par la ligne Residual : 0,036, $p \leq 0{,}001$) que de celui de l'école (0,036 > 0,011).

Tableau 20.3.
Résultats du modèle final – partie aléatoire

Valeurs estimées des paramètres de covariance					
Param de cov	**Sujet**	**Valeur estimée**	**Erreur type**	**Valeur Z**	**Pr > Z**
Intercept	école	0,011	0,004	2,95	0,002
Residual		0,036	0,003	13,13	< 0,000

Le tableau 20.4 rapporte les résultats de la partie fixe du modèle final. L'intervention unique ciblant les habiletés en lecture (*C_INTERV*) n'améliore pas le niveau d'attention des élèves ($B = 0{,}019$[3] ; $t(ddl = 27) = 0{,}34$; n.s.). Toutefois, comme attendu, la proportion de comportements d'attention optimale observée en classe augmente significativement lorsque l'intervention inclut la composante *Attention je lis !* (*C_AJL*) ciblant spécifiquement les comportements d'inattention ($B = 0{,}170$; $t(ddl = 27) = 3{,}58$; $p \leq 0{,}001$). Par ailleurs, le niveau d'attention au prétest (*CON_ATTC*) ($B = -0{,}008$; $t(ddl = 344) = 2{,}71$; $p \leq 0{,}001$) influence négativement la proportion de comportements d'attention optimale observée en classe. Concrètement, cela

3. Le B correspond à la valeur estimée du coefficient de régression et s'interprète comme un bêta.

signifie que des problèmes d'attention au prétest (score élevé sur la variable *CON_ATTC*) sont associés à une proportion plus faible de comportements d'attention optimale observée en classe.

Tableau 20.4.
Résultats du modèle final – partie fixe

Solution pour effets fixes					
Effet	**Valeur estimée**	**Erreur type**	**ddl**	**Valeur du test *t***	**Pr > |t|**
Intercept	0,753	0,048	27	15,97	< 0,000
C_INTERV	0,019	0,055	27	0,34	0,735
C_AJL	0,170	0,047	27	3,58	0,001
CON_ATTC	–0,008	0,003	344	–2,71	0,007

Notes : *C_INTERV* – école de la condition intervention ; *C_AJL* – école de la condition intervention *Attention je lis !* ; CON_ATTC – score d'attention à l'échelle de Conners (2002) remplie par l'enseignant.

Le tableau 20.5 présente le calcul des scores prédits selon la condition et le score d'attention au prétest. Les résultats de ce tableau indiquent que, peu importe la condition expérimentale, les élèves ayant obtenu un score élevé d'attention au prétest ont une proportion de comportements d'attention optimale observée en classe plus élevée que les élèves ayant un niveau d'attention moyen ou faible au prétest. Par ailleurs, alors que les élèves du groupe témoin et ceux de l'intervention unique obtiennent des scores similaires sur la proportion de comportements d'attention optimale observée en classe (respectivement de 71,8 % et de 73,7 % s'il y a, par exemple, présence d'inattention au prétest), les élèves de l'intervention combinée incluant la composante *Attention je lis !* se démarquent par une proportion plus élevée de comportements d'attention optimale en classe (p. ex., 88,8 % pour les élèves qui manifestaient de l'inattention au prétest).

Tableau 20.5.
Calcul des valeurs prédites[a] de la proportion de comportements d'attention optimale observée en classe selon la condition expérimentale et le niveau d'attention au prétest

Condition	Niveau attention[b]	Y_{ij}	=	γ_{00}	+	γ_{01}	$(C_INTERV)_j$	+	γ_{02}	$(C_AJL)_j$	+	γ_{10}	$(CON_ATTC)_{ij}$
Attention, je lis !	Élevé	95,8 %	=	0,753	+	0,019	0	+	0,170	1	+	−0,008	−4,34
	Moyen	92,3 %	=	0,753	+	0,019	0	+	0,170	1	+	−0,008	0
	Inattention	88,8 %	=	0,753	+	0,019	0	+	0,170	1	+	−0,008	4,34
Intervention	Élevé	80,7 %	=	0,753	+	0,019	1	+	0,170	0	+	−0,008	−4,34
	Moyen	77,2 %	=	0,753	+	0,019	1	+	0,170	0	+	−0,008	0
	Inattention	73,7 %	=	0,753	+	0,019	1	+	0,170	0	+	−0,008	4,34
Témoin	Élevé	78,8 %	=	0,753	+	0,019	0	+	0,170	0	+	−0,008	−4,34
	Moyen	75,3 %	=	0,753	+	0,019	0	+	0,170	0	+	−0,008	0
	Inattention	71,8 %	=	0,753	+	0,019	0	+	0,170	0	+	−0,008	4,34

[a] Le calcul des valeurs prédites s'effectue à l'aide d'une équation unifiée, combinant les équations de niveaux 1 et 2 (voir Raudenbush et Bryk, 2002, pour de plus amples détails).

[b] La variable *CON_ATTC* a été centrée autour de la grande moyenne. Un niveau élevé d'attention correspond à un écart-type sous la moyenne (−4,34) ; un niveau moyen correspond à la moyenne (0) ; et l'inattention correspond à un écart-type au-dessus de la moyenne (4,34).

CONCLUSION

En tenant compte de la structure nichée des observations (mesures répétées nichées dans les personnes ; personnes nichées dans des contextes), l'analyse multiniveaux permet d'accommoder la dépendance des observations (Snijders et Bosker, 1999 ; Raudenbush et Bryk, 2002 ; Bickel, 2007). Ce faisant, elle permet, substantiellement, de répondre à des questions qui concernent non seulement l'effet de variables individuelles, mais aussi l'effet de variables de niveaux supérieurs, telles que les caractéristiques d'un milieu de travail. Ce chapitre a fait une présentation sommaire et appliquée du modèle linéaire hiérarchique relevant de l'analyse multiniveaux. Nous y avons examiné des exemples d'études utilisant une structure hiérarchique, les éléments théoriques de base concernant l'analyse multiniveaux, un exemple détaillé d'une analyse multiniveaux, incluant la marche à suivre, la formulation des équations, ainsi que l'interprétation des résultats et les conditions nécessaires justifiant le recours à ce type d'analyse.

RÉFÉRENCES

AL OTAIBA, S. et D. FUCHS (2002). « Characteristics of children who are unresponsive to early literacy intervention : A review of the literature », *Remedial and Special Education*, vol. 23, nº 5, p. 300-316.

ALLARDYCE, J. et J. BOYDELL (2006). « Review : The wider social environment and schizophrenia », *Schizophrenia Bulletin*, vol. 32, nº 4, p. 592-598.

BICKEL, R. (2007). *Multilevel Analysis for Applied Research : It's Just Regression !*, New York, The Guilford Press.

BOYLE, M.H. et J.D. WILLMS (2001). « Multilevel modelling of hierarchical data in developmental studies », *Journal of Child Psychology and Psychiatry and Allied Disciplines*, vol. 42, nº 1, p. 141-162.

BRAULT, M.C. et É. LACOURSE (2012). « Prevalence of prescribed Attention-Deficit Hyperactivity Disorder medications and diagnosis among Canadian preschoolers and school-age children : 1994-2007 », *Canadian Journal of Psychiatry*, vol. 57, nº 2, p. 93-101.

BRAULT, M.C., B. MEULEMAN *et al.* (2012). « Depressive symptoms in the Belgian population : Disentangling age and cohort effects », *Social Psychiatry and Psychiatric Epidemiology*, vol. 47, nº 6, p. 903-915.

BRIÈRE, F.N., S. PASCAL *et al.* (2013). « School environment and adolescent depressive symptoms : A multilevel longitudinal study », *Pediatrics*, vol. 131, nº 3, p. e702-e708.

CONNERS, C. (2000). *Conners Rating Scale-revised*, Toronto, Multi-Health Systems.

DALLY, K. (2006). « The influence of phonological processing and inattentive behavior on reading acquisition », *Journal of Educational Psychology*, vol. 98, nº 2, p. 420.

DIEZ-ROUX, A.V. (1998). « Bringing context back into epidemiology : Variables and fallacies in multilevel analysis », *American Journal of Public Health*, vol. 88, nº 2, p. 216-222.

DIEZ-ROUX, A.V. (2000). « Multilevel analysis in public health research », *Annual Review of Public Health*, vol. 21, nº 1, p. 171-192.

DIEZ-ROUX, A.V. (2002). « A glossary for multilevel analysis », *Journal of Epidemiology and Community Health*, vol. 56, nº 8, p. 588-594.

DION, É., C. BORRI-ANADON *et al*. (2005a). *Apprendre à lire à deux. Manuel de l'enseignante et matériel de lecture*, Montréal, Université du Québec à Montréal.

DION, É., M. POTVIN *et al*. (2005b). *Attention je lis ! Manuel de l'enseignante*, Montréal, Université du Québec à Montréal.

DION, É., C. ROUX *et al*. (2011). « Improving attention and preventing reading difficulties among low-income first-graders : A randomized study », *Prevention Science*, vol. 12, nº 1, p. 70-79.

DUPÉRÉ, V., É. LACOURSE *et al*. (2007). « Méthodes d'analyse du changement fondées sur les trajectoires de développement individuelle : modèles de régression mixtes paramétriques et non paramétriques [1] », *Bulletin de méthodologie sociologique*, vol. 95, nº 1, p. 26-57.

ENDERS, C.K. et D. TOFIGHI (2007). « Centering predictor variables in cross-sectional multilevel models : A new look at an old issue », *Psychological Methods*, vol. 12, nº 2, p. 121.

FARAONE, S.V., R.H. PERLIS *et al*. (2005). « Molecular genetics of attention-deficit/hyperactivity disorder », *Biological Psychiatry*, vol. 57, nº 11, p. 1313-1323.

FOX, W. (1999). *Statistiques sociales*, Québec, Les Presses de l'Université Laval.

HOSENBOCUS, S. et R. CHAHAL (2009). « A review of long-acting medications for ADHD in Canada », *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, vol. 18, nº 4, p. 331.

HOX, J.J. et I.G. KREFT (1994). « Multilevel analysis methods », *Sociological Methods & Research*, vol. 22, nº 3, p. 283-299.

KIDGER, J., R. ARAYA *et al*. (2012). « The effect of the school environment on the emotional health of adolescents : A systematic review », *Pediatrics*, vol. 129, nº 5, p. 925-949.

LECOMTE, T., C. LECLERC, M. CORBIÈRE, T. WYKES, C.J. WALLACE et A. SPIDEL (2008). « Group cognitive behaviour therapy or social skills training for individuals with a first episode of psychosis ? Results of a randomized controlled trial », *Journal of Nervous and Mental Disease*, vol. 196, p. 866-875.

MAAS, C.J. et J.J. HOX (2005). « Sufficient sample sizes for multilevel modeling », *Methodology : European Journal of Research Methods for the Behavioral and Social Sciences*, vol. 1, nº 3, p. 86-92.

PEUGH, J.L. (2010). « A practical guide to multilevel modeling », *Journal of School Psychology*, vol. 48, nº 1, p. 85-112.

PINGAULT, J., S. CÔTÉ *et al.* (2012). « Childhood trajectories of inattention, hyperactivity and oppositional behaviors and prediction of substance abuse/dependence : A 15-year longitudinal population-based study », *Molecular Psychiatry*, vol. 18, p. 806-812.

PINGAULT, J.-B., R.E. TREMBLAY *et al.* (2011). « Childhood trajectories of inattention and hyperactivity and prediction of educational attainment in early adulthood : A 16-year longitudinal population-based study », *American Journal of Psychiatry*, vol. 168, nº 11, p. 1164-1170.

RABINER, D.L. et P.S. MALONE (2004). « The impact of tutoring on early reading achievement for children with and without attention problems », *Journal of Abnormal Child Psychology*, vol. 32, nº 3, p. 273-284.

RAUDENBUSH, S.W. (1993). « A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research », *Journal of Educational and Behavioral Statistics*, vol. 18, nº 4, p. 321-349.

RAUDENBUSH, S.W. et A.S. BRYK (2002). *Hierarchical Linear Models : Applications and Data Analysis Methods*, Thousand Oaks, Sage Publications.

RAUDENBUSH, S. W., A. S. BRYK, Y. F. CHEONG, R. CONGDON et M. DU TOIT (2004). *HLM 6. Linear and nonlinear modeling*, Lincolnwood, Scientific Software International.

SAS INSTITUTE INC. (2002-2010). SAS 9.3, Cary, NC, États-Unis.

SINGER, J. et J. WILLETT, (2003). *Applied Longitudinal Data Analysis : Modeling Change and Event Occurrence*, Oxford, Oxford University Press.

SINGER, J.D. (1998). « Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models », *Journal of Educational and Behavioral Statistics*, vol. 23, nº 4, p. 323-355.

SNIJDERS, T.A. (2005). « Power and sample size in multilevel linear models », dans B. Everitt et D. Howell (dir.), *Encyclopedia of Statistics in Behavioral S*cience, Chichester, Wiley, p. 1570-1573.

SNIJDERS, T.A. et R.J. BOSKER (1999). *Multilevel Analysis : An Introduction to Basic and Advanced Multilevel Modeling*, Thousand Oaks, Sage Publications.

VAN BUUREN, S. (2010). « Multiple imputation of multilevel dat », dans J.J. Hox et J.K. Roberts (dir.), *The Handbook of Advanced Multilevel Analysis*, New York, Routledge, p. 173-196.

YANG, Y. (2007). « Is old age depressing ? Growth trajectories and cohort variations in late-life depression », *Journal of Health and Social Behavior*, vol. 48, nº 1, p. 16.

YANG, Y. et K. LAND (2006). « A mixed models approach to the age-period-cohort analysis of repeated cross-section surveys, with an application to data on trends in verbal test scores », *Sociological Methodology*, vol. 36, nº 1, p. 75-97.

CHAPITRE 21

LA MODÉLISATION PAR ÉQUATIONS STRUCTURELLES
Illustration pour expliquer la motivation à se maintenir en emploi dans une entreprise sociale

Sara Zaniboni
Marc Corbière

FORCES

- Elle offre une structure d'analyse des données qui est large et flexible.
- Elle permet d'évaluer des phénomènes et théories complexes.
- Elle permet d'évaluer des relations entre des variables (observables) et des concepts (construits latents).

LIMITES

- Elle exige une taille d'échantillon conséquente.
- Elle nécessite de prendre soin de certaines conditions relatives aux données avant de réaliser les analyses par équations structurelles (p. ex., données manquantes).
- Lorsque le modèle testé exige de nombreuses modifications, il est nécessaire de tester le nouveau modèle avec un nouvel échantillon.

Les premiers balbutiements de la modélisation par équations structurelles proviennent des travaux de Karl Jöreskog (1973), à partir desquels le premier logiciel intitulé LISREL a été conçu. Grâce aux avantages que cette méthode apporte, la modélisation par équations structurelles (MES)[1] est devenue très populaire et utilisée dans moult disciplines. Deux principales raisons expliquent l'utilisation de la MES. En premier lieu, lorsque comparée aux méthodes et analyses statistiques conventionnelles, la MES a l'avantage de tester des théories ou des modèles théoriques complexes incluant de nombreuses variables dépendantes et indépendantes. La MES représente donc un outil intéressant pour assurer l'avancement de théories, voire d'un champ disciplinaire. La seconde raison invoquée est que la MES peut, d'une part, traiter simultanément des relations complexes entre les variables dépendantes et indépendantes et, d'autre part, vérifier l'assise de la validité d'outils de mesure.

Dans le cadre de ce chapitre, nous proposons dans une première partie de prendre connaissance du vocabulaire de la MES et des conventions qui la définissent (Mueller, 1996 ; Schumacker et Lomax, 2010). Certaines conditions préalables inhérentes aux données seront présentées afin d'expliquer plus en détail les différentes étapes préconisées dans la MES (p. ex., spécification, identification, estimation, évaluation et modification du modèle). Les diverses approches qui peuvent être utilisées pour évaluer des modèles structurels complexes (p. ex., désagrégation totale ou partielle, agrégation totale ou partielle) seront également abordées dans cette première partie (Schumacker et Lomax, 2010 ; Ullman, 2007). Enfin, une illustration de la MES sera proposée dans une seconde partie du chapitre pour suivre de façon pragmatique l'utilisation et l'interprétation des différents éléments qui la composent. L'illustration de la MES porte sur la vérification d'un modèle théorique permettant d'expliquer la motivation à se maintenir en emploi de personnes aux prises avec un trouble mental grave (p. ex., schizophrénie, trouble de l'humeur) travaillant dans une entreprise sociale localisée en Italie.

1. DÉFINITION, VOCABULAIRE ET CONVENTIONS DE LA MES

La modélisation par équations structurelles (*structural equation modeling*) est une méthode statistique multivariée avancée qui permet de tester de façon empirique des modèles théoriques complexes (Jöreskog, 1973).

1. Dans ce chapitre, nous avons opté pour le vocable ou l'expression *modélisation par équations structurelles*, mais il est possible de rencontrer dans la littérature spécialisée l'expression *modèle(s) d'équation(s) structurelle(s)/structurale(s)*.

Plus précisément, en partant d'un modèle théorique, le chercheur sera en mesure de tester, grâce à la MES, comment un ensemble de variables définissent des construits ou concepts, et comment ces variables et concepts s'articulent entre eux. On comprend alors que l'un des objectifs principaux de l'utilisation de la MES reste de vérifier dans quelle mesure un modèle théorique est appuyé par des données brutes recueillies dans un projet de recherche. Différents types de modèles théoriques peuvent être testés par la MES, dont diverses analyses statistiques (p. ex., analyses de régression, pistes causales, analyse factorielle confirmatoire). Dans la mesure où les données n'expliquent pas le modèle théorique qui est testé, le chercheur peut alors modifier le modèle original ou concevoir de nouveaux modèles théoriques afin de les tester à nouveau (Byrne, 1998). Autrement dit, d'un mode confirmatoire, il est possible de passer à un mode heuristique; quoique certains auteurs s'en défendront et préconiseront exclusivement le mode confirmatoire de la MES.

Dans l'objectif de bien comprendre la MES, il est important de s'approprier son vocabulaire, en définissant certaines notions essentielles ainsi qu'en présentant quelques modèles de base (Mueller, 1996; Schumacker et Lomax, 2010). Tout d'abord, il est à noter qu'il y a une distinction claire entre une variable latente et une variable que l'on observe (ou observable): la première (latente) est représentée dans les graphiques de la MES par un ovale ou une ellipse, alors que la seconde (observable) est représentée par un rectangle ou un carré. Les variables latentes ne peuvent pas être directement observées ou mesurées; elles peuvent l'être seulement indirectement ou inférées à partir des observables. En ce qui concerne les variables dites observables, elles sont la plupart du temps des réponses (ou scores) à des énoncés d'un questionnaire.

Une autre distinction importante est celle qui se situe entre les variables dépendantes ou variables *exogènes* et les variables dites indépendantes, appelées aussi variables *endogènes*. Les variables endogènes ne sont pas influencées par d'autres variables dans le modèle théorique à tester, et, dans les représentations graphiques de la MES, aucune flèche ne les pointe (Mueller, 1996). Par contre, les variables exogènes sont des variables qui sont influencées par d'autres variables, et dans les graphiques de la MES, les flèches sont pointées sur elles (Schumacker et Lomax, 2010).

Ces précisions lexiques et graphiques nous permettent ainsi de présenter les modèles de base inclus dans la MES. Par exemple, l'analyse de régression (voir le chapitre 18 de cet ouvrage) est un modèle ou une analyse statistique qui inclut seulement des observables et une seule variable dépendante peut être prédite par une ou plusieurs variables indépendantes. Les pistes causales renvoient quant à elles à un modèle ou à une analyse

statistique qui comprend seulement des observables, mais qui peut inclure plusieurs variables dépendantes et indépendantes, d'où leur plus grande complexité lorsque comparées à l'analyse de régression. Une analyse factorielle confirmatoire (voir le chapitre 22 de cet ouvrage) renvoie à un modèle statistique incluant des observables qui permettent d'évaluer des variables latentes, qu'elles soient dépendantes ou indépendantes. La MES est en fait une combinaison de ces modèles de base susmentionnés, notamment les pistes causales (c.-à-d. un modèle structurel) et un modèle factoriel confirmatoire (c.-à-d. un modèle de mesure) (Byrne, 1998). Pour finir à propos des notions incontournables à considérer, les variables latentes et observables, les variables exogènes et endogènes ainsi que l'erreur de mesure associée à ces deux types de variables sont incluses dans la MES.

1.1. Triage des données (*data screening*)

Sachant que la MES est basée sur une méthode de recherche corrélationnelle, une première étape consiste à vérifier les données en vue de repérer d'éventuels problèmes qui pourraient rendre difficile la réalisation de la MES (Kline, 2011; Schumacker et Lomax, 2010). Les aspects qui peuvent avoir des répercussions sur les matrices de variance-covariance sont le niveau et le type de mesure, une étendue restreinte des valeurs associées aux données, des données manquantes, la non-normalité des données, les données aberrantes ou extrêmes (*outliers*), la multicolinéarité des données et les problèmes de taille de l'échantillon. L'ensemble de ces aspects sera abordé dans les prochaines lignes.

Les *propriétés des échelles de mesure* (p. ex., nominale, ordinale, à intervalle et continue) sont reliées à différents types de coefficients de corrélation. Chacune de ces échelles de mesure pourrait être éventuellement utilisée dans la MES, toutefois, il est recommandé de ne pas les mettre ensemble dans une matrice de corrélation ou de covariance. En général, comme pour les régressions linéaires, les pistes causales et les analyses factorielles, la MES exige des variables mesurées sur un type de mesure à intervalle ou continue. Ces deux types de variables (à intervalle ou continue) devraient aussi comprendre une étendue des scores suffisamment grande pour assurer une variance des données. En fait, si l'étendue des scores est restreinte, la magnitude de la valeur des corrélations sera également réduite.

Les *données manquantes* présentent un autre obstacle, car elles peuvent influencer les résultats. Ces premières peuvent être repérées grâce à une analyse des données manquantes. Le type de décision pour gérer les

données manquantes (p. ex., retranchement des sujets qui présentent des données manquantes telles que les méthodes *listwise* ou *pairwise*, substitution par la moyenne, régression ou maximum de vraisemblance) ont des répercussions sur la magnitude, la direction des corrélations et du nombre de sujets retenus pour l'analyse.

La *normalité ou la non-normalité des données* : en général, la MES assume que les observables sont des données multivariées normalement distribuées. Les tests d'asymétrie (*skewness*) et d'aplatissement des données (*kurtosis*) sont aussi disponibles dans la plupart des logiciels statistiques pour prendre connaissance de la normalité (ou non) des données. À titre d'exemple, une méthode classique pour traiter la non-normalité des données est l'utilisation de la matrice de covariance asymptotique.

La présence de *données aberrantes ou extrêmes* (*outliers*) est un autre problème possible qui doit être pris au sérieux avant de commencer la MES. Ces données aberrantes ou extrêmes peuvent se produire pour diverses raisons (p. ex., erreurs lors de la saisie des données, erreurs dues à l'outil de mesure) et ont donc des conséquences sur les moyennes, l'écart-type et les coefficients de corrélation.

La *multicolinéarité* a lieu lorsque les variables sont initialement considérées comme distinctes alors que dans la réalité, ces variables évaluent des choses semblables. Ces difficultés peuvent être surmontées grâce à des procédures disponibles dans les logiciels statistiques (p. ex., le calcul de la corrélation au carré – ou variance entre chacune des variables –, ou les critères d'inflation de la variance [*variance inflation factor* –VIF]).

La *taille de l'échantillon* a besoin d'être prise en considération lorsqu'on souhaite tester un modèle théorique grâce à la MES. Cependant, il n'y a pas de consensus sur la taille d'échantillon recommandée pour la MES. En général, une grande taille d'échantillon est exigée pour atteindre la puissance de l'analyse requise et pour obtenir une estimation stable des paramètres à évaluer. Des modèles plus complexes exigent aussi des échantillons plus conséquents. Une règle tacite veut qu'habituellement la taille d'échantillon soit supérieure à 200 sujets pour tester un modèle théorique ou comprenne entre 5 et 20 fois plus de sujets que le nombre de paramètres à estimer. Par ailleurs, il est convenu que si le modèle a déjà été testé dans d'autres études, alors la taille de l'échantillon peut être moindre que la règle tacite mentionnée ci-dessus (Kline, 2011).

1.2. Étapes de la MES

En général, on compte cinq étapes principales que le chercheur devrait respecter lors de l'utilisation de la MES : 1) la spécification du modèle ; 2) l'identification du modèle ; 3) l'estimation du modèle ; 4) l'évaluation du modèle ; et 5) la modification du modèle. Ces cinq étapes sont présentées ci-dessous avec plus de détails (Schumacker et Lomax, 2010 ; Ullman, 2007).

1.2.1. Spécification du modèle

Avant même d'entreprendre la collecte des données, le chercheur élabore un modèle théorique qu'il souhaite tester, et ce, en considérant toutes les théories pertinentes dans la littérature scientifique ainsi que toutes les informations utiles à la conception du modèle théorique (p. ex., résultats d'études). Autrement dit, le chercheur doit spécifier un modèle pour le confirmer ou l'infirmer avec les données qu'il souhaite recueillir. Lors de cette phase, le chercheur doit aussi décider quelles variables à inclure dans le modèle et définir comment elles seront reliées entre elles. Lorsque le modèle théorique qui a été conçu n'est pas identique au modèle à tester, il est alors incorrectement spécifié. Une erreur dans la spécification du modèle peut se glisser quand certaines variables pertinentes ont été écartées du modèle théorique original.

1.2.2. Identification du modèle

La question de l'identification du modèle reste complexe (pour une discussion à ce sujet, consulter le livre de Kline, 2011). À cause de l'aspect flexible de la spécification du modèle, plusieurs modèles peuvent être conçus. Cependant, ce ne sont pas tous les modèles qui sont *spécifiés* et qui peuvent être *identifiés* ou *estimés*. L'idée de base est qu'il ne devrait y avoir qu'un seul ensemble de paramètres à estimer, lesquels sont compris dans la matrice de corrélation/covariance à évaluer. Chaque paramètre dans la MES doit être *spécifié*, c'est-à-dire *libre* (à noter qu'un paramètre qui est libre vise à être évalué), *fixé* (un paramètre qui n'est pas *libre* est *fixé* à une valeur qui, habituellement, est de 0 ou 1), ou encore *contraint* (un paramètre qui est *contraint* se doit d'être égal à un ou plusieurs autres paramètres). Une fois que les paramètres sont spécifiés *fixés*, *libres* ou *contraints*, ils sont mis ensemble pour former un seul et unique modèle associé à une matrice particulière. Si tous les paramètres du modèle sont identifiés, alors le modèle en entier est identifié, c'est-à-dire que tous les paramètres de la MES peuvent être estimés. Quand le modèle ne peut pas être identifié, les résultats ne peuvent pas être interprétés, et le modèle nécessite une nouvelle spécification. On compte trois niveaux d'identification de modèle : *sur-identifié*, *juste-identifié*

et *sous-identifié*. Un modèle est dit *sur-identifié* quand il comprend plus de variables que de paramètres à estimer. Un modèle est dit *juste-identifié* quand il contient le même nombre de variables que de paramètres à estimer. Un modèle est dit *sous-identifié* quand il compte moins de variables que de paramètres à estimer. Lorsqu'un modèle est *sous-identifié*, il peut être considéré comme non identifié, alors que lorsqu'il est *juste-identifié* ou *sur-identifié*, il peut être considéré comme *identifié* (Schumacker et Lomax, 2010). Il existe certaines conditions pour établir l'identification d'un modèle, et il est préconisé d'utiliser différentes méthodes afin d'éviter des problèmes d'identification[2].

1.2.3. Estimation du modèle

Un modèle correctement spécifié a des paramètres (p. ex., paramètres libres) qui peuvent être estimés à partir des données disponibles. L'objectif premier du processus d'estimation est d'évaluer les paramètres à travers des méthodes qui minimisent l'écart entre les variables observées ou observables de la matrice de covariance (les données) et le modèle théorique que l'on souhaite tester (modèle). Différentes méthodes telles que le maximum de vraisemblance[3] et d'autres (p. ex., Generalized Least Squares – GLS, Unweighted or Ordinary Least Squares – ULS ou OLS) peuvent être utilisées afin d'estimer les paramètres du modèle théorique (Schumacker et Lomax, 2010). L'estimation d'un modèle pourrait ne pas converger[4] ou encore, les solutions fournies par le logiciel pourraient être erronées. En l'occurrence, un message peut être fourni par le logiciel dans la sortie de résultats (*output*), comme « les estimés des paramètres ne peuvent pas être interprétés » ou encore « les coefficients sont en dehors de l'étendue souhaitée (entre – 1 et 1) ».

1.2.4. Évaluation du modèle

L'évaluation du modèle consiste à tester si le modèle théorique est soutenu par les données recueillies. Autrement dit, l'évaluation permet de savoir si les données empiriques (ou brutes) convergent avec le modèle théorique testé. Deux aspects importants nécessitent d'être pris en compte : 1) évaluer l'ajustement total du modèle en fonction des indices d'ajustement (*fit indices*) ; 2) évaluer les estimations des paramètres (Raykov et Marcoulides, 2006).

---

2. Par exemple, des modèles non récursifs (c.-à-d. qui ne sont pas en boucle) permettent de contourner ce problème.
3. C'est la méthode d'estimation qui est la plus fréquemment utilisée et qui se trouve par défaut dans la plupart des logiciels pour réaliser la MES.
4. En général, les logiciels MES interrompent le processus d'estimation et envoient un message d'erreur.

*Évaluation des indices d'ajustement (*fit indices*)*

Plusieurs indices d'ajustement existent, et il est suggéré d'évaluer le modèle en considérant plusieurs d'entre eux, plutôt que de se restreindre à un seul indice (Schumacker et Lomax, 2010). Pour obtenir plus de détails quant à l'utilisation de ces indices d'ajustement et des discussions qui en découlent, nous suggérons au lecteur de consulter plusieurs ouvrages, notamment ceux auxquels sont associés Bentler (Bentler et Raykov, 2000 ; Hu et Bentler, 1999). Les indices d'ajustement peuvent être divisés en trois grandes catégories : 1) l'ajustement du modèle ; 2) la comparaison du modèle ; et 3) la parcimonie du modèle.

Ajustement du modèle

Il détermine le degré auquel les données de l'échantillon s'ajustent à la MES établie *a priori*. Les indices d'ajustement sont fondés sur les différences entre la matrice observée (celle relative aux données recueillies) et la matrice de variance-covariance (celle relative au modèle à tester). Dans cette catégorie sont inclus le chi-carré (χ^2), le chi-carré normé ou le ratio chi-carré sur le nombre de degrés de liberté (Normed Chi-square – NC ou χ^2/ddl), l'erreur de mesure au carré approximative (Root Mean Square Error of approximation – RMSEA), l'indice de bon ajustement (Goodness-of-Fit Index – GFI) et l'indice de bon ajustement ajusté (Adjusted Goodness-of-Fit Index – AGFI).

Le chi-carré est la mesure classique pour évaluer le bon ajustement du modèle aux données recueillies et permet d'évaluer la magnitude de l'écart entre les données de l'échantillon et l'ajustement des matrices de covariances (Hu et Bentler, 1999). Une valeur non significative du χ^2 signifie qu'il n'y a pas de différence significative entre la matrice de l'échantillon et celle que l'on souhaite prédire avec le modèle posé. Une faible valeur du χ^2, avec des niveaux de signification plus grands que 0,05 ($p > 0{,}05$), indique qu'il n'y a pas de différences significatives entre les deux matrices (données brutes et modèle théorique à tester), et donc, on peut conclure qu'il y a un bon ajustement des données brutes au modèle théorique testé (Barrett, 2007). Autrement dit, lorsqu'on exécute une MES, on vise que la valeur du chi-carré ne soit pas significative (contrairement à d'autres analyses classiques). Bien que le test du chi-carré soit encore très populaire pour ce type de statistique, il présente de nombreuses limites. Tout d'abord, le χ^2 est sensible à la taille de l'échantillon, notamment lorsque la taille augmente (p. ex., un échantillon plus grand que 200 individus), aussi, il a tendance à montrer un seuil de probabilité significatif ($p < 0{,}05$), qui signifie que le modèle est rejeté (Bentler et Bonett, 1980). Par contre, lorsque les échantillons sont plus petits (p. ex., un échantillon inférieur à 100 individus), le test du chi-carré manque de puissance et a tendance à montrer un niveau

de probabilité non significatif, qui signifie que le test ne discrimine pas entre un bon ajustement et un moins bon pour la MES à tester (Kenny et McCoach, 2003). Par conséquent, lorsqu'on utilise des échantillons trop petits ou trop grands pour tester un modèle théorique, le test du chi-carré demeure un test à utiliser avec précaution. La deuxième limite du test du chi-carré vient du fait qu'il est aussi sensible aux écarts avec la normalité multivariée des variables observées (ou observables) et en retour, ce même test peut rejeter le modèle qui est testé, et ce, même quand le modèle est spécifié adéquatement (McIntosh, 2007).

En résumé, même si le test du chi-carré présente de nombreuses limites, il devrait toujours être rapporté lors de la présentation des résultats. À cause des limites associées au test du chi-carré, une évaluation alternative est d'observer le ratio du chi-carré et le nombre de degrés de liberté (ddl) du modèle, soit le chi-carré normé (NC ou χ^2/ddl) (Jöreskog, 1969). Bien qu'il n'y ait pas de consensus à propos du ratio acceptable dans le cadre de la MES, une petite valeur du χ^2/ddl indique un bon ajustement du modèle aux données recueillies. Il y a cependant quelques recommandations par les auteurs qui suggèrent que l'étendue acceptable devrait se situer entre 1 et 5, avec comme indication qu'une valeur inférieure à 1 renvoie à un faible ajustement des données au modèle théorique alors qu'une valeur supérieure à 5 indique que le modèle théorique exige d'être révisé (Schumacker et Lomax, 2010)[5].

L'indice RMSEA mesure l'écart entre la matrice de covariance estimée (modèle théorique) et celle qui est brute (observables) par le nombre de degrés de liberté (Steiger, 1990). Cet indice est reconnu comme l'un des critères les plus informatifs de la structure de la matrice de covariance modélisée. L'indice RMSEA est en faveur de la parcimonie et sa valeur décroît s'il y a plus de degrés de liberté (plus grande parcimonie) ou si un échantillon de grande taille est considéré. Les valeurs sont sur un continuum de 0 à 1, où 0 indique le meilleur ajustement. En général, les valeurs inférieures à 0,05 indiquent un bon ajustement, les valeurs de 0,05 à 0,08 un ajustement acceptable, celles entre 0,08 et 0,10 indiquent un ajustement médiocre et celles qui sont supérieures à 0,10 montrent un très faible ajustement au modèle théorique.

L'indice GFI a été créé comme une alternative aux limites du test du chi-carré et mesure la proportion de variance et de covariance dans la matrice observée et expliquée par le modèle. L'AGFI ajuste le GFI qui est

---

5. Une valeur qui oscille entre 2 et 4 serait donc considérée comme acceptable, avec la valeur la plus faible comme la plus adéquate. Ces repères restent toutefois indicatifs et ne peuvent pas être considérés comme des normes.

basé sur les degrés de liberté. L'étendue des valeurs de ces deux indices oscille entre 0 et 1, où 1 indique le meilleur ajustement. Il est suggéré que les valeurs des indices GFI et AGFI de 0,90 ou au-dessus peuvent représenter une bonne approximation des données au modèle testé (Hu et Bentler, 1999). Une limite des indices GFI et AGFI vient de ce que leurs valeurs varient selon la taille de l'échantillon. Plus précisément, lorsque de grands échantillons sont utilisés, la valeur des indices a tendance à augmenter.

Comparaison du modèle

Les indices de la comparaison du modèle permettent de comparer le modèle proposé (fondé sur la théorie) avec un modèle indépendant (aussi connu comme le modèle nul) qui représente le cas extrême, soit aucune relation parmi les variables observées. Le modèle nul est relatif à la valeur du chi-carré du modèle indépendant. Dans cette catégorie sont inclus l'indice d'ajustement normé (Normed Fit Index – NFI), l'indice d'ajustement non normé (Non-Normed Fit Index – NNFI) et l'indice d'ajustement comparatif (Comparative Fit Index – CFI).

Les indices NFI et NNFI sont conçus pour comparer le modèle proposé avec le modèle dans lequel aucune relation n'est assumée pour l'ensemble des variables. Le NFI (Bentler et Bonett, 1980) est calculé en mettant en relation la valeur du chi-carré pour le modèle proposé à la valeur du chi-carré du modèle nul. La principale limite de cet indice est qu'il est sensible à la taille de l'échantillon et ainsi, l'indice d'ajustement est sous-estimé pour des échantillons inférieurs à 200 individus (Bentler, 1990). L'indice NNFI est une variante de l'indice NFI qui prend en compte les degrés de liberté dans le modèle proposé. L'indice NNFI semble plus robuste face aux variations de la taille de l'échantillon et, par conséquent, son évaluation est recommandée dans les études. Cependant, à cause de sa nature non normée, l'indice NNFI peut prendre des valeurs qui vont au-delà de 1, d'où la difficulté quand on l'interprète (Byrne, 1998). L'indice NFI (ainsi que le NNFI en général) présente des étendues de valeurs de 0 à 1, avec une valeur égale ou supérieure à 0,90 indiquant un ajustement acceptable. Hu et Bentler (1999) soutiennent que le seuil acceptable pour ces indices devrait être égal ou supérieur à 0,95.

L'indice CFI (Bentler, 1990) est une forme révisée de l'indice NFI, qui présente une certaine efficacité même dans le cas de petits échantillons. Cet indice évalue l'ajustement relatif à d'autres modèles dont l'approche est basée sur la distribution non centrale du chi-carré (*the noncentral* χ^2 *distribution*). Comme pour les indices NFI et NNFI, l'indice CFI a une étendue des

valeurs oscillant de 0 à 1, avec une valeur égale ou supérieure à 0,90 qui indique un ajustement acceptable. Cependant, un seuil révisé, proche de 0,95, a été proposé (Hu et Bentler, 1999).

Parcimonie du modèle

Les indices du modèle de parcimonie prennent en compte la complexité du modèle dans l'évaluation de l'ajustement du modèle, notamment le nombre de paramètres exigés à estimer pour atteindre une valeur de chi-carré donnée. Cette catégorie inclut l'indice du bon ajustement de la parcimonie (Parsimony Goodness-of-Fit Index – PGFI), l'indice d'ajustement normé parcimonieux (Parsimonious Normed Fit Index – PNFI) et le critère d'information d'Akaike (Akaike Information Criterion – AIC).

Mulaik *et al*. (1989) ont conçu les indices PGFI et PNFI. L'indice PGFI est basé sur l'indice GFI, tout en étant ajusté aux degrés de liberté. L'indice PNFI s'ajuste également aux indices de liberté, mais il est toutefois basé sur l'indice NNFI. Ces indices, PGFI et PNFI, tiennent compte du nombre de paramètres à estimer et, par conséquent, les modèles plus complexes sont lésés lors de leur évaluation. Les valeurs de ces indices varient de 0 à 1, où 1 indique le meilleur ajustement. Il n'y a pas de seuil précis recommandé pour ces indices, toutefois, des valeurs qui se situent autour de 0,50 sont plutôt courantes (Byrne, 1998).

Le critère AIC (Akaike, 1987) est généralement utilisé quand des modèles non nichés (voir le chapitre 20 de cet ouvrage) sont comparés. Ainsi, ce critère indique au chercheur quel est le modèle le plus parcimonieux, car ce dernier doit être préféré à celui qui est plus complexe. Aucun seuil précis n'est recommandé pour cet indice. Toutefois, les valeurs les plus faibles indiquent un meilleur ajustement aux données et donc le modèle qui présente le plus faible coefficient AIC est le modèle à retenir.

Évaluation des estimations des paramètres

Une fois que le modèle s'ajuste adéquatement aux données empiriques, les estimations pour chaque paramètre peuvent être interprétées. Plusieurs aspects principaux peuvent être considérés. Le premier est que chaque estimation du paramètre devrait faire sens. En effet, certains résultats pourraient être incorrects et le chercheur devrait alors examiner attentivement toute magnitude inattendue (p. ex., les corrélations ne devraient pas être supérieures à 1) ou la direction des estimations des paramètres (p. ex., les variances ne devraient pas présenter des valeurs négatives), ou encore d'importantes estimations d'erreurs de mesure standard lorsque comparées avec d'autres qui proviennent de la même échelle. Une fois que les estimations

des paramètres sont obtenues, les erreurs de mesure pour chaque estimation sont calculées. Un ratio des estimations des paramètres aux estimations des erreurs de mesure forme une valeur critique (souvent appelée valeur du *t*). Si la valeur critique (*critical ratio* – CR) est plus grande que 1,96 (en valeur absolue), cela renvoie à une valeur significativement différente ($p < 0,05$). Plus précisément, le ratio critique (*two-tailed tests of significance of t-statistic*) peut être interprété comme suit : CR = 1,64 alors $p < 0,10$; CR = 1,96 alors $p < 0,05$; CR = 2,58 alors $p < 0,01$; CR = 3,29 alors $p < 0,001$. Un troisième aspect à vérifier est si le signe du paramètre à estimer (positif ou négatif) correspond à ce qui est attendu (ou pas) pour le modèle théorique.

1.2.5. Modification du modèle

Lorsque le modèle hypothétique est rejeté sur la base des indices d'ajustement, les chercheurs sont souvent intéressés à trouver un modèle alternatif qui s'ajuste aux données empiriques. En l'occurrence, il s'agit de modifier le modèle théorique et ultérieurement d'évaluer ce nouveau modèle. Pour déterminer comment modifier le modèle, les chercheurs peuvent utiliser la recherche de spécification (*specification search* [Leamer, 1978]), qui vise à modifier le modèle original par la recherche d'un modèle qui s'ajuste mieux aux données brutes. En outre, il existe deux procédures, le multiplicateur de Lagrange et le test de Wald, lesquels permettent de déceler les erreurs de spécification.

Le multiplicateur de Lagrange (*Lagrange multiplier*) considère la signification statistique des paramètres à estimer dans le modèle et vise à « réparer » ceux qui peuvent contribuer significativement au modèle. Pour ce faire, le multiplicateur de Lagrange suggère d'évaluer ou *de rendre libres* (voir plus haut la notion de paramètre libre) de nouveaux paramètres (p. ex., une corrélation entre deux variables) qui n'ont pas été spécifiés dans le modèle, mais qui présentent un niveau de signification non négligeable pour le modèle. Pour ce qui est du test de Wald (*Wald test*), il s'agit de la procédure inverse pour laquelle il est suggéré de ne pas évaluer des paramètres qui étaient initialement considérés comme significatifs dans le modèle testé (p. ex., la relation entre deux concepts). Dit autrement, le test de Wald permet de déterminer quel paramètre du modèle pourrait être rejeté afin d'améliorer ce dernier et le rendre plus parcimonieux. Bien qu'aucune règle tacite n'existe sur l'interprétation de ces indices pour considérer une modification significative du modèle, en général, les indices de modification plus grands que 3,84 mériteraient une plus grande attention. Comme la recherche de spécification est de nature exploratoire, les fondements théoriques doivent également être pris en considération. Retenir une amélioration indiquée par le test de Wald ou du multiplicateur de Lagrange est approprié seulement

si la modification du modèle est théoriquement soutenue. La réplication du modèle avec des échantillons indépendants est importante, plus particulièrement lorsque les modifications du modèle sont nombreuses et ont été apportées *post hoc*.

1.3. Diverses approches

Plusieurs approches pour évaluer la MES sont recommandées dans la littérature scientifique (Bagozzi et Heatherton, 1994), telles que la désagrégation totale ou partielle et l'agrégation totale ou partielle.

La *désagrégation totale* utilise les réponses de chaque item (p. ex., item d'un questionnaire) comme les indicateurs distincts d'un construit latent. Cette procédure fournit un niveau d'analyse détaillé pour l'évaluation de la MES. Cependant, l'approche de la désagrégation totale est très sensible aux erreurs de mesure, ce qui rend plus difficile l'obtention d'un modèle satisfaisant en termes d'ajustement aux données brutes.

L'*agrégation totale* représente quant à elle un seul indicateur pour tous les items mesurant le construit ou le concept latent; ces indicateurs consistent souvent en l'agrégation des scores d'une échelle (p. ex., score global d'une échelle). Cette approche fournit des modèles simples et offre aux chercheurs la capacité d'évaluer l'essence sous-jacente d'un concept. Cependant, l'agrégation totale devrait être utilisée seulement dans des situations où les items d'un construit partagent assez de variance commune et où les propriétés uniques des items du construit ne sont pas exigées (contrairement au modèle de mesure).

L'*agrégation partielle* renvoie à l'agrégation d'items de chaque dimension du concept, où chaque dimension sous-jacente est retenue. Un indicateur composé (c.-à-d. un agrégat d'items connus comme groupe d'items [*item parcels*]) est créé à partir des items de chaque dimension distincte du construit, telle qu'un score d'une facette ou d'une sous-dimension d'un questionnaire. Cette approche offre la possibilité d'avoir un contenu substantiel pour chaque variable à l'intérieur d'une plus petite matrice, et ce, en évitant un plus grand nombre d'erreurs de mesure. Cependant, il est recommandé que ces scores composites soient créés à partir d'échelles ou sous-échelles pour lesquelles la structure présente une seule dimension et dont la cohérence interne est bien établie.

La *désagrégation partielle* fournit deux à trois indicateurs composites pour chaque construit à évaluer. Les scores des indicateurs sont créés à partir des réponses aux items, d'où les groupes d'items pourraient être conçus par une agrégation au hasard. La désagrégation partielle permet l'usage d'un

nombre d'indicateurs plus grand pour représenter un construit, et permet ainsi de réduire le nombre de paramètres à estimer (cela permet aussi de contourner les exigences de la taille d'échantillon si trop de variables sont suggérées). Dans le cadre de l'approche de la désagrégation partielle, Bagozzi et Heatherton (1994) suggèrent entre cinq et sept items qui peuvent être divisés en deux indicateurs. S'il y a plus de neuf items, alors cela pourrait être trois indicateurs ou plus.

2. ILLUSTRATION DE LA MES : L'ARTICULATION DE THÉORIES POUR EXPLIQUER LA MOTIVATION DE PERSONNES À SE MAINTENIR EN EMPLOI

Dans le cadre de cette partie, l'illustration consiste à utiliser la MES pour vérifier un modèle théorique relatif à la motivation à se maintenir en emploi de personnes aux prises avec un trouble mental grave (p. ex., schizophrénie). Un nombre inestimable d'articles scientifiques portent sur cette thématique, car le maintien en emploi s'avère la pierre d'achoppement de la réintégration au travail de personnes aux prises avec un trouble mental grave (p. ex., Lanctôt, Corbière et Durand, 2012). Pour ce faire, le logiciel LISREL a été utilisé pour entreprendre les analyses de la MES. Nous nous emploierons à suivre les conventions et étapes de la MES présentées dans la première partie de ce chapitre.

2.1. Triage des données

Les données ont été recueillies auprès de 310 personnes avec un trouble mental grave travaillant dans une entreprise d'économie sociale italienne. L'ensemble des questionnaires administrés permettait d'évaluer quatre concepts relatifs au modèle théorique à tester (voir ci-dessous). Un premier questionnaire composé de douze items (échelle de réponse de 1 = tout à fait en désaccord à 7 = tout à fait d'accord) permettait d'évaluer le soutien social reçu en dehors du milieu de travail (SSOut), provenant de la famille, des amis et des autres personnes significatives (Zimet *et al.*, 1988). Pour évaluer le soutien social provenant du milieu de travail (SSWork), notamment du supérieur immédiat et des collègues de travail (Karasek *et al.*, 1998), un questionnaire de onze items (échelle de réponse en cinq points, 1= pas du tout à 5 = tout à fait) a été utilisé. Un troisième questionnaire composé de neuf énoncés (échelle de réponse de 0 = jamais à 6 = tous les jours) a permis d'évaluer l'engagement au travail (WE) (Balducci, Fraccaroli et Schaufeli, 2010; Schaufeli, Bakker et Salanova, 2006). En vue d'évaluer la motivation à maintenir son emploi (MKW) (Villotti *et al.*, soumis), nous avons utilisé un questionnaire de sept items (échelle de réponse de 1 = tout à fait en désaccord

à 7 = tout à fait en accord). Quand cela s'avérait nécessaire, les items des échelles étaient traduits en italien en utilisant l'approche de la traduction inversée (voir le chapitre 24 de cet ouvrage). Aux fins de cette illustration, nous avons utilisé une désagrégation partielle des items, c'est-à-dire que pour chaque concept évalué ci-dessus, nous avons créé trois groupes d'items (entre deux et quatre items pour chaque groupe d'items).

Avant de réaliser les analyses de la MES, nous avons vérifié si quelques problèmes statistiques pouvaient être inhérents aux données recueillies (p. ex., non-normalité multivariée, multicolinéarité, données manquantes). Comme la normalité multivariée n'était pas totalement respectée, nous avons utilisé une matrice de covariance asymptotique comme point de départ (voir première partie du chapitre), ainsi que la matrice de covariance de l'échantillon.

2.2. Étapes des analyses pour la MES

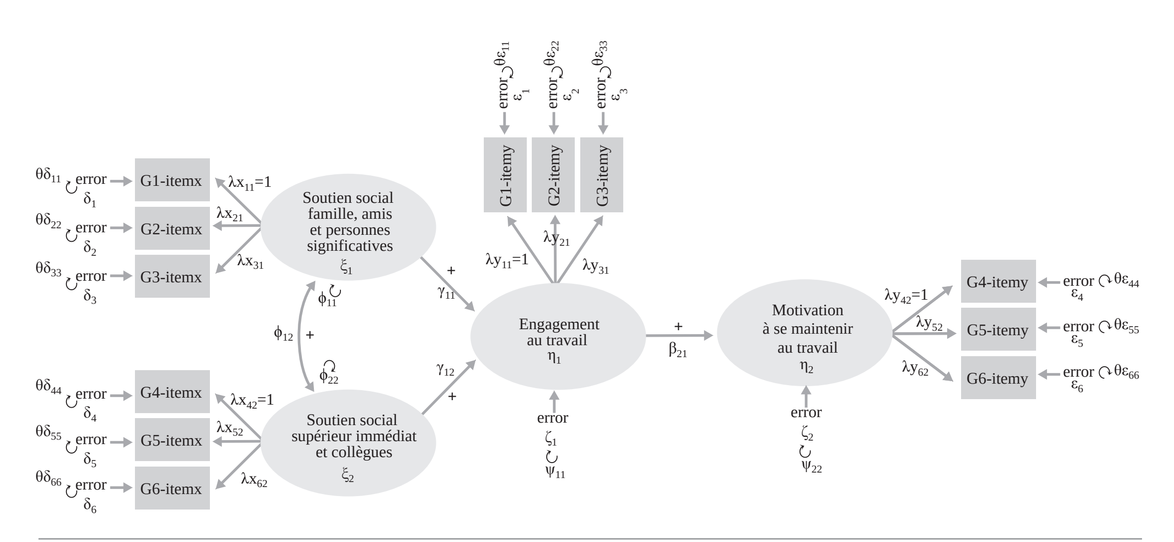
2.2.1. Spécification du modèle

La spécification du modèle exige de considérer les résultats d'études antérieures et des théories pertinentes pour construire le modèle théorique à tester, lequel devra converger avec la structure de la matrice de variance-covariance de l'échantillon. Dans le cadre de cette illustration, le modèle théorique d'intérêt (figure 21.1) renvoie à la prédiction de la motivation à se maintenir en emploi de personnes aux prises avec un trouble mental grave, travaillant dans une entreprise d'économie sociale. Les éléments conceptuels de la littérature scientifique montrent que les notions de soutiens naturels et d'accommodements de travail (Corbière *et al.*, accepté), de motivation et d'engagement au travail (Corbière *et al.*, 2011 ; Villotti *et al.*, 2013) sont importantes à considérer pour prédire les résultats de réintégration au travail des personnes avec un trouble mental grave. En ce sens, le modèle théorique présenté à la figure 21.1 illustre les relations entre l'ensemble de ces concepts, les flèches indiquant la direction de l'influence d'un concept sur un autre.

La MES présente quatre variables latentes (ellipses) : deux sont des variables indépendantes, soit le soutien social offert dans le milieu de travail et le soutien social offert à l'extérieur du milieu de travail. Deux variables sont dépendantes, c'est-à-dire la motivation à se maintenir au travail et l'engagement au travail. Une flèche qui rejoint une variable latente (indépendante ou dépendante) à une autre variable latente symbolise qu'un coefficient doit être estimé. Dans la MES à tester, on note trois coefficients que nous voulons estimer pour ces variables latentes. Aussi, deux variables latentes

indépendantes sont corrélées (double flèche). Qui plus est, chaque variable latente dépendante comprend plusieurs coefficients ainsi qu'une prédiction d'erreur unique qui indique qu'une certaine portion de la variance de la variable n'est pas prédite par les variables latentes indépendantes. Le modèle de mesure (*measurement model*) a trois indicateurs ou observables (rectangles) pour chaque variable latente. Des flèches partent de chacune des variables latentes vers les indicateurs (observables) qui lui sont assignés,

Figure 21.1.
Modèle théorique à tester, incluant les symboles du logiciel LISREL



Note : Le logiciel LISREL utilise les matrices pour spécifier les modèles. Les matrices et commandes sont écrites avec deux lettres de spécification :
LX (lambda x, Λx) fait référence à la matrice de coefficients de saturation entre les variables indépendantes, observables et latentes (λx) ;
LY (lambda y, Λy) fait référence à la matrice de coefficients de saturation entre les variables dépendantes, observables et latentes (λy) ;
GA (gamma, Γ) fait référence à la matrice des coefficients de régression d'une variable latente dépendante prédite par une variable latente indépendante (γ) ;
BE (bêta, β) fait référence à la matrice des coefficients de régression d'une variable dépendante latente qui prédit une autre variable dépendante latente (β) ;
PH (phi, Φ) fait référence à la matrice de variances-covariances parmi les variables latentes indépendantes (φ) ;
PS (psi, Ψ) fait référence à une matrice de variances-covariances des erreurs associées aux variables latentes dépendantes (ζ) ;
TD (theta-delta, θ_δ) fait référence à la matrice des variances-covariances parmi les erreurs associées aux variables indépendantes observables (δ) ;
TE (theta-epsilon, θ_ε) fait référence à la matrice des variances-covariances parmi les erreurs associées aux variables dépendantes observables (ε).
Pour de plus amples informations concernant ces matrices et symboles, le lecteur est invité à consulter le logiciel de Jöreskog et Sörbom (1996).

et pour chaque indicateur ou observable, un coefficient de saturation au concept est calculé (voir le chapitre 22 de cet ouvrage). Pour chaque observable (ou groupe d'items en l'occurrence), il y a une évaluation de l'erreur de mesure qui est indiquée par une flèche qui pointe l'observable. Autrement dit, cela signifie qu'une portion de chaque observable est mesurée par quelque chose d'autre qui n'est pas formulé sous le format d'une hypothèse. Dans l'annexe 21.1, la syntaxe pour conduire la MES de cette illustration est présentée.

2.2.2. Identification du modèle

Une fois que le modèle théorique est spécifié, la prochaine étape est l'identification du modèle. L'identification du modèle permet de décider si un ensemble unique de paramètres à estimer peut être calculé. Dans la MES présentée à la figure 21.1, certains paramètres sont fixés alors que d'autres sont libres, pour une estimation de la matrice de variance-covariance de l'échantillon étudié. Un exemple de paramètre fixé est le premier indicateur de chaque groupe d'items (ou observables) appartenant à la variable latente, dont le coefficient de saturation est égal à 1 (p. ex., λx_{11} à la figure 21.1). Un exemple de paramètre libre est celui qui rejoint les deux variables latentes, le soutien social dans le milieu de travail et l'engagement au travail (p. ex., γ_{12} à la figure 21.1). Enfin, après vérification avec le logiciel LISREL, le modèle proposé est identifié (voir la section 1.2.2).

2.2.3. Estimation du modèle

La prochaine étape consiste à estimer les paramètres au sein de la MES à tester. Comme mentionné plus haut, les paramètres peuvent être estimés selon diverses procédures, et nous avons choisi de retenir celle qui permet de contrôler la non-normalité des données, soit le maximum de vraisemblance robuste (Robust Maximum Likelihood – RML).

2.2.4. Évaluation du modèle

*Évaluation des indices d'ajustement (*fit indices*)*

Si l'on considère les indices d'ajustement de la sortie de résultats 1 de LISREL (p. ex., les indices en gras de la figure 21.2), la MES testée semble converger avec les données brutes (voir la section 1.2.4 pour comparer).

Figure 21.2.
Sortie des résultats 1 – LISREL

Goodness of Fit Statistics

Degrees of Freedom = 50
Minimum Fit Function Chi-Square = 121.74 (P = 0.00)
Normal Theory Weighted Least Squares Chi-Square = 119.15 (P = 0.00)
Satorra-Bentler Scaled Chi-Square = 98.50 (P = 0.00)
Chi-Square Corrected for Non-Normality = 114.88 (P = 0.00)
Estimated Non-centrality Parameter (NCP) = 48.50
90 Percent Confidence Interval for NCP = (24.12 ; 80.68)

Minimum Fit Function Value = 0.39
Population Discrepancy Function Value (F0) = 0.16
90 Percent Confidence Interval for F0 = (0.078 ; 0.26)
Root Mean Square Error of Approximation (RMSEA) = 0.056
90 Percent Confidence Interval for RMSEA = (0.040 ; 0.072)
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.26

Expected Cross-Validation Index (ECVI) = 0.50
90 Percent Confidence Interval for ECVI = (0.42 ; 0.60)
ECVI for Saturated Model = 0.50
ECVI for Independence Model = 10.55

Chi-Square for Independence Model with 66 Degrees of Freedom = 3236.79
Independence AIC = 3260.79
Model AIC = 154.50
Saturated AIC = 156.00
Independence CAIC = 3317.63
Model CAIC = 287.13
Saturated CAIC = 525.45

Normed Fit Index (NFI) = 0.97
Non-Normed Fit Index (NNFI) = 0.98
Parsimony Normed Fit Index (PNFI) = 0.73
Comparative Fit Index (CFI) = 0.98
Incremental Fit Index (IFI) = 0.98
Relative Fit Index (RFI) = 0.96

Critical N (CN) = 239.90

Root Mean Square Residual (RMR) = 1.32
Standardized RMR = 0.073
Goodness of Fit Index (GFI) = 0.94
Adjusted Goodness of Fit Index (AGFI) = 0.91
Parsimony Goodness of Fit Index (PGFI) = 0.60

Évaluation des estimations des paramètres

En se référant (figure 21.3) à la sortie des résultats 2 de LISREL, les estimations des paramètres sont statistiquement significatives au seuil $p < 0{,}01$ (*critical ratio*/*t-value* = 2,58), et le signe positif (+) est en accord avec ce qui était attendu des relations entre les concepts du modèle théorique (voir les signes des flèches dans la figure 21.1). Par exemple, le degré de motivation à maintenir son emploi est influencé par l'engagement au travail qui, en retour, est influencé par le soutien social reçu par le milieu de travail et par le soutien social reçu à l'extérieur du travail.

Figure 21.3.
Sortie des résultats 2 – LISREL

LISREL Estimates (Robust Maximum Likelihood)

LAMBDA-Y*

	WE	MKJ	
G1-itemy	1.00	– –	
G2-itemy	1.02	– –	parameter
	(0.05)		standard error
	18.72		**critical ratio/t-value**
G3-itemy	1.00	– –	
	(0.05)		
	20.20		
G4-itemy	– –	1.00	
G5-itemy	– –	0.61	
		(0.04)	
		15.87	
G6-itemy	– –	0.53	
		(0.04)	
		13.42	

LAMBDA-X

	SSOut	SSWork
G1-itemx	1.00	– –
G2-itemx	1.20	– –
	(0.13)	
	9.00	
G3-itemx	1.20	– –
	(0.12)	
	9.81	
G4-itemx	– –	1.00
G5-itemx	– –	0.57
		(0.06)
		8.89

(*suite*)

G6-itemx —— 0.77		
		(0.07)
		11.35

BETA

	WE	MKJ	
WE	——	——	
MKJ	0.52	——	parameter
	(0.08)		standard error
	6.70		**critical ratio/t-value**

GAMMA

	SSOut	SSWork
WE	0.25	0.39
	(0.08)	(0.13)
	3.08	**2.92**
MKJ	——	——

PHI

	SSOut	SSWork
SSOut	25.89	
	(4.65)	
	5.56	
SSWork	10.04	10.27
	(1.70)	(1.83)
	5.90	**5.60**

PSI
Note : cette matrice est diagonale.

WE	MKJ
11.48	11.45
(1.57)	(1.58)
7.33	**7.25**

* Comme dans la figure 21.1, lambda fait référence à la matrice de coefficients de saturation entre les variables dépendantes, observables et latentes du modèle testé.

La section de la sortie des résultats 3 de LISREL (figure 21.4) intitulée *Squared multiple correlations* offre les informations relatives au pourcentage de variance qui est expliquée pour chacune des variables dépendantes à partir des autres variables qui la pointent par une flèche (autres variables dépendantes ou indépendantes). Dans la sortie des résultats 3 de LISREL, on peut observer que le modèle explique 31 % de la variance (R^2 = 0,31) de l'engagement au travail et 28 % de la variance (R^2 = 0,28) de la motivation à se maintenir au travail.

Figure 21.4.
Sortie des résultats 3 – LISREL

Squared Multiple Correlations for Structural Equations

WE	MKJ
--------	--------
0.31	0.28

Dans la sortie des résultats 4 de LISREL (figure 21.5), les erreurs associées aux variables exogènes (theta-epsilon) et celles associées aux variables endogènes (theta-delta) sont rapportées. Les corrélations multiples au carré (*squared multiple correlations*) pour les variables Y font référence au pourcentage de la variance de ces indicateurs exogènes attribuée aux variables latentes exogènes plutôt qu'à l'erreur de mesure.

Figure 21.5.
Sortie des résultats 4 – LISREL

THETA-EPS

G1-itemy	G2-itemy	G3-itemy	G4-itemy	G5-itemy	G6-itemy
6.44	6.36	4.13	3.29	1.46	2.60
(0.92)	(1.15)	(0.76)	(0.89)	(0.32)	(0.40)
6.98	5.51	5.45	3.71	4.57	6.56

Squared Multiple Correlations for Y – Variables

G1-itemy	G2-itemy	G3-itemy	G4-itemy	G5-itemy	G6-itemy
0.72	0.73	0.80	0.83	0.80	0.63

THETA-DELTA

G1-itemx	G2-itemx	G3-itemx	G4-itemx	G5-itemx	G6-itemx
41.60	33.85	20.55	12.45	3.65	3.96
(4.82)	(4.56)	(3.25)	(1.57)	(0.52)	(0.62)
8.62	7.42	6.32	7.91	7.07	6.39

Squared Multiple Correlations for X – Variables

G1-itemx	G2-itemx	G3-itemx	G4-itemx	G5-itemx	G6-itemx
0.38	0.52	0.64	0.45	0.48	0.61

Les estimations standardisées de la solution finale sont rapportées dans la sortie des résultats 5 de LISREL (figure 21.6). La solution est acceptable et satisfaisante étant donné que les estimations standardisées ne sont pas supérieures à 1 en valeur absolue.

Figure 21.6.
Sortie des résultats 5 – LISREL

Completely Standardized Solution

LAMBDA-Y

	WE	MKJ
G1-itemy	0.85	– –
G2-itemy	0.86	– –
G3-itemy	0.90	– –
G4-itemy	– –	0.91
G5-itemy	– –	0.89
G6-itemy	– –	0.80

LAMBDA-X

	SSOut	SSWork
G1-itemx	0.62	– –
G2-itemx	0.72	– –
G3-itemx	0.80	– –
G4-itemx	– –	0.67
G5-itemx	– –	0.69
G6-itemx	– –	0.78

BETA

	WE	MKJ
WE	– –	– –
MKJ	0.53	– –

GAMMA

	SSOut	SSWork
WE	0.32	0.31
MKJ	– –	– –

Correlation Matrix of ETA and KSI

	WE	MKJ	SSOut	SSWork
WE	1.00			
MKJ	0.53	1.00		
SSOut	0.51	0.27	1.00	
SSWork	0.50	0.27	0.62	1.00

PSI

Note : cette matrice est diagonale.

WE	MKJ
0.69	0.72

THETA-EPS

G1-itemy	G2-itemy	G3-itemy	G4-itemy	G5-itemy	G6-itemy
0.28	0.27	0.20	0.17	0.20	0.37

THETA-DELTA

G1-itemx	G2-itemx	G3-itemx	G4-itemx	G5-itemx	G6-itemx
0.48	0.36	0.55	0.52	0.39	

2.2.5. Modification du modèle

Lorsque nous considérons la sortie des résultats 6 de LISREL (figure 21.7), nous constatons que dans la matrice gamma, l'index de modification, notamment pour le lien entre le soutien social du milieu de travail et la motivation à se maintenir au travail – SSWork et MKW (c.-à-d. γ_{22}) –, est significativement grand et qu'il y aurait lieu de considérer ce lien additionnel dans le modèle afin de rendre ce modèle plus solide ou plus convergent avec les données brutes. Ce lien a été ajouté à la figure 21.8[6], car il présente sans équivoque une pertinence sur le plan théorique (Corbière *et al.*, 2011). Il en résulte que le modèle révisé présenté à la figure 21.8 (ajout du lien entre soutien social en milieu de travail et motivation à se maintenir au travail) a des indices d'ajustement supérieurs au modèle initial (voir la figure 21.1). Ce modèle final est retenu aux fins de l'illustration.

6. La figure 21.8 est un bon exemple de figure qui devrait être présentée dans un article scientifique.

Figure 21.7.
Sortie des résultats 6 – LISREL

Modification Indices and Expected Change

...

Modification Indices for GAMMA

	SSOut	SSWork
WE	– –	– –
MKJ	8.62	**57.31**

Expected Change for GAMMA

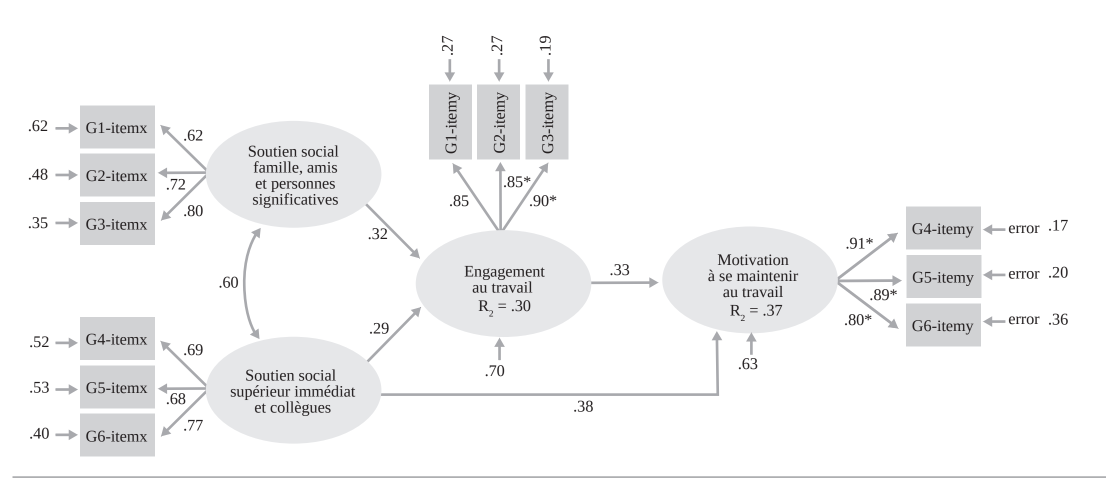
	SSOut	SSWork
WE	– –	– –
MKJ	0.22	0.96

Standardized Expected Change for GAMMA

SSOut SSWork

WE	– –	– –
MKJ	0.28	0.77

Figure 21.8.
Modèle révisé



Notes : n = 310 ; * p < 0,01 ; les paramètres de la solution standardisée finale et complète sont présentés ; le modèle révisé présente un bon ajustement aux données brutes avec les indices d'ajustement suivants : χ^2 (49) = 91,72, p < 0,05 ; χ^2/ddl = 1,87 ; RMSEA = 0,04 ; NNFI = 0,99 ; CFI = 0,99.

CONCLUSION

La modélisation par équations structurelles (MES) est une méthode rigoureuse qui permet aux chercheurs en sciences humaines, sociales et de la santé de tester des modèles théoriques ou des liens conceptuels à partir de données recueillies dans un projet de recherche. La MES est sans conteste utile pour le développement de théories, mais elle offre aussi aux chercheurs une meilleure compréhension de l'articulation des liens conceptuels de variables exogènes et endogènes. Les logiciels pour exécuter des MES sont nombreux et parfois difficiles à maîtriser. Un investissement en temps et en énergie est nécessaire pour se familiariser avec la MES et plus particulièrement avec le logiciel LISREL, qui a son propre lexique. Dans le cadre de ce chapitre, nous avons tenté d'ailleurs de rendre les plus explicites possibles les étapes à franchir pour les analyses relatives à la MES et à l'interprétation des résultats qui en découlent. Au-delà de l'utilisation de ces logiciels, il est fortement recommandé de suivre les formations requises pour mieux comprendre les mécanismes sous-jacents à la MES.

Annexe 21.1.
Syntaxe de LISREL

```
da  ni=12  no=310  ma=cm

cm  fi=MKW.cov
ac  fi=MKW.acm

la
Parcely1  Parcely2  Parcely3  Parcely4  Parcely5  Parcely6  Parcelx1  Parcelx2
Parcelx3  Parcelx4  Parcelx5  Parcelx6

se
1  2  3  4  5  6  7  8  9  10  11  12/

mo  ny=6  ne=2  nx=6  nk=2  ps=sy,fi  ph=sy,fr  ly=fu,fi  lx=fu,fi  te=di,fr  td=di,fr
ga=fu,fi  be=fu,fi

le
WE  MKJ

lk
SSOut  SSWork

fr  ly  2  1  ly  3  1  ly  5  2  ly  6  2
fr  lx  2  1  lx  3  1  lx  5  2  lx  6  2
fr  ga  1  1  ga  1  2
fr  be  2  1
fr  ps  1  1  ps  2  2
va  1  ly  1  1  ly  4  2  lx  1  1  lx  4  2

path  diagram
ou  rs  sc  mi  ad=off
```

RÉFÉRENCES

AKAIKE, H. (1987). « Factor analysis and AIC », *Psychometrika*, vol. 52, nº 3, p. 317-332, <doi : 10.1007/bf02294359>.

BAGOZZI, R.P. et T.F. HEATHERTON (1994). « A general approach to representing multifaceted personality constructs : Application to state self-esteem », *Structural Equation Modeling : A Multidisciplinary Journal*, vol. 1, nº 1, p. 35-67, <doi : 10.1080/10705519409539961>.

BALDUCCI, C., F. FRACCAROLI et W.B. SCHAUFELI (2010). « Psychometric properties of the Italian version of the Utrecht Work Engagement Scale (UWES-9) : A cross-cultural analysis », *European Journal of Psychological Assessment*, vol. 26, nº 2, p. 143-149, <doi : 10.1027/1015-5759/A000020>.

BARRETT, P. (2007). « Structural equation modelling : Adjudging model fit », *Personality and Individual Differences*, vol. 42, nº 5, p. 815-824, <doi : 10.1016/j.paid.2006.09.018>.

BENTLER, P.M. (1990). « Comparative fit indexes in structural models », *Psychological Bulletin*, vol. 107, nº 2, p. 238-246.

BENTLER, P.M. et D.G. BONETT (1980). « Significance tests and goodness of fit in the analysis of covariance structures », *Psychological Bulletin*, vol. 88, nº 3, p. 588-606, <doi : 10.1037/0033-2909.88.3.588>.

BENTLER, P.M. et T. RAYKOV (2000). « On measures of explained variance in nonrecursive structural equation models », *Journal of Applied Psychology*, vol. 85, nº 1, p. 125-131, <doi : 10.1037/0021-9010.85.1.125>.

BYRNE, B.M. (1998). *Structural Equation Modeling with LISREL, PRELIS, and SIMPLIS : Basic Concepts, Applications, and Programming*, Mahwah, Lawrence Erlbaum Associates Publishers.

CORBIÈRE, M., P. VILLOTTI, T. LECOMTE, G. BOND, A. LESAGE et E. GOLDNER (sous presse). « Work accommodations and natural supports for maintening employment », *Psychiatric Rehabilitation Journal*.

CORBIÈRE, M., S. ZANIBONI, T. LECOMTE, G.R. BOND, P.Y. GILLES, A. LESAGE et E. GOLDNER (2011). « Job acquisition for people with severe mental illness enrolled in supported employment programs : A theoretically grounded empirical study », *Journal of Occupational Rehabilitation*, vol. 21, nº 3, p. 342-354, <doi : 10.1007/s10926-011-9315-3>.

HU, L.T. et P.M. BENTLER (1999). « Cutoff criteria for fit indexes in covariance structure analysis : Conventional criteria versus new alternatives », *Structural Equation Modeling*, vol. 6, nº 1, p. 1-55, <doi : 10.1080/10705519909540118>.

JÖRESKOG, K.G. (1969). « A general approach to confirmatory maximum likelihood factor analysis », *Psychometrika*, vol. 34, nº 2 (Pt.1), p. 183-202, <doi : 10.1007/bf02289343>.

JÖRESKOG, K.G. (1973). « A general method for estimating a linear structural equation system », dans A.S. Goldberger et O.D. Duncan (dir.), *Structural Equation Models in the Social Sciences*, New York, Seminar, p. 85-112.

JÖRESKOG, K.G. et D. SÖRBOM (1996). *LISREL 8 : User's Reference Guide*, Lincolnwood, Scientific Software International.

Karasek, R., C. Brisson, N. Kawakami, I. Houtman, P. Bongers et B. Amick (1998). « The Job Content Questionnaire (JCQ) : An instrument for internationally comparative assessments of psychosocial job characteristics », *Journal of Occupational Health Psychology*, vol. 3, nº 4, p. 322-355.

Kenny, D.A. et D.B. McCoach (2003). « Effect of the number of variables on measures of fit in structural equation modeling », *Structural Equation Modeling*, vol. 10, nº 3, p. 333-351, <doi : 10.1207/s15328007sem1003_1>.

Kline, R.B. (2011). *Principles and Practice of Structural Equation Modeling*, 3e éd., New York, Guilford Press.

Lanctôt, N., M. Corbière et M.-J. Durand (2012). « Job tenure and quality of work life of people with psychiatric disabilities working in social enterprises », *Journal of Vocational Rehabilitation*, vol. 37, nº 1, p. 39-48.

Leamer, E.E. (1978). *Specification Searches : Ad Hoc Inference with Nonexperimental Data*, New York, Wiley.

McIntosh, C.N. (2007). « Rethinking fit assessment in structural equation modelling : A commentary and elaboration on Barrett », *Personality and Individual Differences*, vol. 42, nº 5, p. 859-867, <doi : 10.1016/j.paid.2006.09.020>.

Mueller, R.O. (1996). *Basic Principles of Structural Equation Modeling : An Introduction to LISREL and EQS*, New York, Springer.

Mulaik, S.A., L.R. James, J. Van Alstine, N. Bennett, S. Lind et C.D. Stilwell (1989). « Evaluation of goodness-of-fit indices for structural equation models », *Psychological Bulletin*, vol. 105, nº 3, p. 430-445, <doi : 10.1037/0033-2909.105.3.430>.

Raykov, T., et G. Marcoulides. (2006). *A First Course in Structural Equation Modeling*, 2e éd., NJ :, Lawrence Erlbaum.

Schaufeli, W.B., A.B. Bakker et M. Salanova (2006). « The measurement of work engagement with a short questionnaire : A cross-national study », *Educational and Psychological Measurement*, vol. 66, nº 4, p. 701-716, <doi : 10.1177/0013164405282471>.

Schumacker, R.E. et R.G. Lomax (2010). *A Beginner's Guide to Structural Equation Modeling*, 3e éd., New York, Routledge.

Steiger, J.H. (1990). « Structural model evaluation and modification : An interval estimation approach », *Multivariate Behavioral Research*, vol. 25, nº 2, p. 173-180, <doi : 10.1207/s15327906mbr2502_4>.

Ullman, J.B. (2007). « Structural equation modeling », dans B.G. Tabachnick et L.S. Fidell (dir.), *Using Multivariate Statistics*, 5e éd., Boston, Allyn and Bacon, p. 676-780.

Villotti, P., C. Balducci, S. Zaniboni, M. Corbière et F. Fraccaroli (2014). « An analysis of work engagement among workers with mental disorders recently integrated to work », *Journal of Career Assessment*, vol. 22, nº 1, p. 18-27.

Villotti, P., M. Corbière, S. Zaniboni, T. Lecomte et F. Fraccaroli (soumis). « Evaluating the motivation to sustain employment in people with severe mental illness working in social enterprises ».

Zimet, G.D., N.W. Dahlem, S.G. Zimet et G.K. Farley (1988). « The multidimensional scale of perceived social support », *Journal of Personality Assessment*, vol. 52, nº 1, p. 30-41, <doi : 10.1207/s15327752jpa5201_2>.



LES ANALYSES FACTORIELLES EXPLORATOIRES ET CONFIRMATOIRES
Illustration à l'aide de données recueillies sur l'estime de soi en tant que travailleur

Marc Corbière

FORCES

- L'analyse factorielle exploratoire permet de regrouper des observables (p. ex., variables) homogènes en des concepts ou dimensions conceptuelles.
- L'analyse factorielle confirmatoire permet de confirmer ou d'infirmer le regroupement d'observables sous un même concept qui est développé *a priori*.
- L'analyse factorielle (exploratoire et confirmatoire) vise la parcimonie.

LIMITES

- Elles peuvent mener à des solutions factorielles qui peuvent différer selon l'échantillon qui est évalué.
- Les résultats dépendent des connaissances théoriques et empiriques de ceux qui entreprennent l'analyse factorielle, notamment celles relatives aux concepts à évaluer.
- L'analyse factorielle exploratoire peut entraîner des biais, lesquels peuvent survenir à la suite de la prise de décision concernant le nombre de facteurs à extraire.

Au début du siècle dernier, Spearman (1904) a inventé l'analyse factorielle. Cette analyse renvoie à un modèle explicatif corrélationnel qui permet de regrouper, sous un ou plusieurs facteurs, un ensemble d'observables homogènes (c.-à-d. indicateurs, variables, énoncés, items). Autrement dit, à partir d'un bassin de données ou d'observables, l'analyse factorielle permet d'extraire un ou plusieurs facteurs (ou dimensions conceptuelles) au sein desquels les observables sont regroupés et significativement corrélés. Par ailleurs, les facteurs émergeant de l'analyse factorielle peuvent aussi présenter une intercorrélation.

Il existe une différence notable entre l'analyse factorielle exploratoire et celle de type confirmatoire. Comme leur adjectif les qualifie, la première analyse factorielle n'impose aucune solution factorielle préétablie, alors que la seconde permet de confirmer ou d'infirmer une solution factorielle qui est mise à l'épreuve. On comprend alors qu'au moins dans le cas de l'analyse factorielle confirmatoire, un fondement théorique ou empirique est requis pour inférer un patron de relations entre des observables appartenant à un facteur et, ainsi, tester ces hypothèses ou inférences.

Dans le cadre de ce chapitre, une première partie sera consacrée à la présentation de l'analyse factorielle de type exploratoire et ensuite à celle de type confirmatoire. La partie relative à l'analyse factorielle confirmatoire sera succincte puisqu'elle est un cas particulier de la modélisation par équations structurelles présentée dans un autre chapitre de ce livre (voir le chapitre 21). Dans la partie suivante, une application concrète de ces deux types d'analyses factorielles sera illustrée grâce à des données recueillies auprès de personnes aux prises avec un trouble mental grave (p. ex., schizophrénie) en processus de recherche d'emploi. Dans le cadre de ce chapitre, il s'agira de valider l'outil intitulé Estime de soi en tant que travailleur auprès de ce public (Corbière *et al.*, 2009).

1. INTRODUCTION À L'ANALYSE FACTORIELLE EXPLORATOIRE

L'analyse factorielle exploratoire (AFE) consiste en plusieurs objectifs non exclusifs tels que le développement de construits ou de concepts, la validité d'échelles ou d'outils de mesure (p. ex., questionnaire), la réduction du nombre d'observables par facteur(s) en vue de refléter l'essence d'un construit, appelée aussi la loi de la parcimonie (Thompson, 2004). La loi de la parcimonie postule que lorsque deux explications correspondent à un ensemble de faits quasiment semblables, la plus simple est la plus vraisemblable. Les chercheurs s'efforcent d'ailleurs d'expliquer le plus de variance commune pour rendre compte d'un phénomène à l'étude, en utilisant le moins d'observables et de facteurs possible (Henson et Roberts, 2006). Dans

les prochains paragraphes, les notions statistiques et critères à considérer dans l'utilisation et l'interprétation d'une AFE seront présentés selon quatre étapes : 1) avant l'utilisation de l'AFE ; 2) lors de l'utilisation d'une AFE initiale ; 3) au moment de fixer le nombre de facteurs à extraire de l'AFE initiale ; et 4) après avoir fixé le nombre de facteurs à extraire, une nouvelle AFE est réalisée.

1.1. Éléments à considérer avant l'utilisation d'une AFE

Dans cette section, la taille de l'échantillon, le type d'observables requis, la variance commune, les nuances entre l'analyse par composantes principales (ACP) et l'analyse factorielle exploratoire (AFE) seront abordés. Par ailleurs, comme c'est le cas pour d'autres analyses multivariées, il convient de tenir compte des données manquantes ou aberrantes (*outliers*), de la multicolinéarité, de la non-violation de la normalité des observables avant même l'utilisation des analyses factorielles (Tabachnick et Fidell, 2007). Ces dernières précautions ne seront toutefois pas abordées dans cette section, car elles sont couvertes par d'autres chapitres de ce livre (p. ex., le chapitre 18).

En ce qui concerne la *taille de l'échantillon* requise pour l'analyse factorielle, Thompson (2004) mentionne qu'un ratio observable/individus de 1 pour 10 est le critère sur lequel la plupart des chercheurs s'entendent. D'autres auteurs (Bryant et Yarnold, 1995 ; Gorsuch, 1983 ; Tabachnick et Fidell, 2001) suggèrent toutefois qu'un ratio de 1 pour 5 est le minimum requis, avec un échantillon qui ne doit jamais être inférieur à 100 individus. Enfin, d'autres stipulent que l'échantillon ne devrait jamais être inférieur à 300 sujets (Norušis et SPSS Inc., 2005 ; Tabachnick et Fidell, 2001). Eu égard à ces critères qui varient d'un auteur à un autre, Beavers *et al.* (2013), dans leur recension des écrits, se sont penchés sur la question et observent qu'au-delà du nombre minimal d'individus et du ratio observable/sujets, d'autres critères devraient être pris en compte, notamment le nombre d'observables appartenant à chaque facteur et les indices de saturation de ces observables aux facteurs (voir explications sur cette notion plus bas). Ils précisent entre autres que si, d'une part, on considère 10 à 12 items (ou observables) avec des indices de saturation aux facteurs de l'ordre de 0,40 ou plus, alors un échantillon de 150 individus est suffisant. D'autre part, si les facteurs sont définis avec peu d'observables et présentent des indices de saturation faibles à modérés, alors un échantillon de 300 sujets est requis. Par son mode exploratoire, on comprend qu'il est difficile de déterminer *a priori* combien de facteurs vont émerger de l'analyse et quels seront les indices de saturation des observables aux facteurs. À cet effet, Beavers *et al.* (2013) concluent que puisque l'AFE fait partie de la famille des méthodes

multivariées, un échantillon d'au moins 150 sujets pour l'AFE devrait être un prérequis pour des analyses factorielles initiales. Au-dessous de cette taille d'échantillon (N = 150), les auteurs devront justifier les raisons qui y sont sous-jacentes (p. ex., difficulté de recrutement des individus) et être prudents quant à l'interprétation des résultats.

Pour conduire une AFE, les observables sont plus souvent évalués sur une *échelle à intervalle ou continue*, comme l'échelle de Likert. Dans le cadre de projets de recherche, les observables peuvent faire référence à des énoncés (ou items) d'un questionnaire. Lorsque la matrice de corrélation est analysée, il est fortement recommandé d'observer des coefficients de corrélation égaux ou supérieurs à 0,30 pour l'ensemble des observables (Tabachnick et Fidell, 2007). Sans cette condition, les données ne devraient pas être utilisées pour une AFE, ou du moins, le chercheur devrait anticiper certaines difficultés pour l'interprétation des résultats obtenus. Par ailleurs, il n'est pas non plus souhaitable d'observer des coefficients de corrélation très élevés ($r > 0{,}70$) qui indiqueraient une possible multicolinéarité entre les observables.

L'AFE et l'analyse par composantes principales (ACP) sont toutes deux considérées comme des analyses dont l'objectif est de réduire le nombre d'observables à un nombre de facteurs restreint. Bien que ces deux types d'analyses soient similaires sur plusieurs aspects et utilisées de façon interchangeable par les chercheurs, elles présentent cependant leurs propres caractéristiques. La factorisation par axes principaux (*principal axis factoring*, PAF) – nommée communément l'AFE – est utilisée pour expliquer la *variance commune*[1] des observables, c'est-à-dire la variance partagée par ces dernières. Elle se distingue, en ce sens, de l'ACP. L'ACP cherche, quant à elle, à extraire la variance totale, c'est-à-dire les variances commune, unique et reliée à l'erreur de mesure (Hair *et al.*, 2006). En d'autres mots, l'AFE est un modèle qui décrit ce qui est commun aux observables (la variance commune) alors que l'ACP décrit la variance totale des observables (sans aucune distinction entre les types de variance). Malgré cette distinction importante, il est à noter que l'ACP et l'AFE sont souvent confondues dans les articles publiés dans les revues scientifiques. Il ne faut donc pas tenir pour acquis que lorsque dans un article scientifique, les auteurs mentionnent qu'ils ont utilisé une AFE, il s'agit bien d'une AFE (cela pourrait être une ACP). Cette confusion[2] peut aussi s'expliquer par le fait que techniquement parlant,

1. La *variance commune* est la variance d'une variable ou d'un observable qui est partagée avec d'autres variables incluses dans l'analyse. La *variance unique* et la *variance relative à l'erreur de mesure* sont toutes deux propres à une variable en particulier (voir aussi le glossaire dans O'Brien, 2007).
2. Le lecteur est invité à consulter d'autres articles concernant la distinction entre ces deux types d'analyses (p. ex., Fabrigar *et al.*, 1999; Schmitt, 2011).

l'ACP est réalisée à partir d'une matrice de corrélations comportant des unités sur la diagonale et c'est pourquoi on la considère parfois comme une analyse factorielle particulière, c'est-à-dire une « analyse factorielle avec des unités (valeur de 1) sur la diagonale » (Dassa, 1996, p. 63). Tabachnick et Fidell (2001, p. 610-611) stipulent que si le chercheur est intéressé par une solution théorique qui n'est pas *contaminée* par les variances uniques et associées à l'erreur, l'AFE est le meilleur choix. Si, par contre, le chercheur souhaite obtenir un portrait de l'ensemble des observables, l'ACP reste alors l'analyse à retenir. La variance totale des observables peut présenter une certaine pertinence pour le chercheur qui souhaite appréhender le phénomène étudié dans toute son ampleur. C'est pourquoi certains auteurs, à juste titre, sont d'avis que l'ACP a tendance à gonfler la variance totale des résultats obtenus ainsi que les indices de saturation des observables aux facteurs (Costello et Osborne, 2005 ; Fabrigar *et al.*, 1999). Bien que l'ACP présente cette limite et que l'AFE permette d'obtenir des informations plus précises sur la variance commune, partagée entre les observables, les résultats demeurent sensiblement similaires, et ce, d'autant plus que les observables évalués sont valides et fidèles (Beavers *et al.*, 2013). Enfin, Pett, Lackey et Sullivan (2003) suggèrent de comparer les résultats de l'ACP et de l'AFE (c.-à-d. PAF) afin de déterminer la meilleure explication ou solution factorielle possible, tout en considérant l'assise théorique du phénomène étudié. C'est ce qui sera d'ailleurs proposé dans l'illustration de ce chapitre.

1.2. Éléments à considérer lors de l'utilisation d'une AFE initiale

Parmi les critères à considérer lors de l'utilisation et de l'interprétation des résultats d'une AFE initiale, on retient le Bartlett's Test of Sphericity et le Kaiser-Meyer-Olkin Measure of sampling adequacy (KMO) (Thompson, 2004).

Le Bartlett's Test of Sphericity[3] (Bartlett, 1954) permet d'évaluer si la matrice de corrélation est une matrice de type *identité* (*identity matrix*). Une telle matrice consiste à avoir des 1 sur la principale diagonale et des 0 au-dessus et au-dessous de cette diagonale. Si l'échantillon est supérieur à 500 individus, le test de Bartlett aura tendance à être significatif pour des coefficients de corrélation faibles ($r < 0{,}20$). Le seuil de signification pour ce test est d'une valeur de $p < 0{,}05$.

3. La notion de sphéricité signifie que les données ou observables ne sont pas intercorrélés.

Le KMO (Kaiser, 1970, 1974) consiste en la corrélation partielle entre deux observables en contrôlant pour les autres observables. Kaiser (1974) suggère le barème ci-dessous pour donner un aperçu des seuils acceptables du KMO. En outre, le coefficient KMO se situe entre 0 et 1 et il devrait être supérieur à 0,60 (Tabachnick et Fidell, 2001) :

0,90 = coefficient excellent ;

0,80 = coefficient très satisfaisant ;

0,70 = coefficient moyennement satisfaisant ;

0,60 = coefficient médiocre ;

0,50 = coefficient proche de l'échec ;

< 0,50 = coefficient inacceptable.

1.3. Éléments à considérer pour déterminer le nombre de facteurs à extraire de l'AFE initiale

Pour déterminer avec plus de précision le nombre de facteurs à extraire de l'AFE initiale, plusieurs critères sont à considérer, dont la valeur propre (*eigenvalue*) des facteurs, le test d'éboulis (*scree plot*), la variance cumulée pour l'ensemble des facteurs et l'analyse parallèle (*parallel analysis*) (Cattell, 1966 ; Guttmann, 1954 ; Thompson, 2004).

La *valeur propre* (*eigenvalue*) représente la somme de variance qui est associée à un facteur. Il faut savoir que la somme des valeurs propres est égale au nombre total de variables ou d'observables. On comprend alors que la valeur propre d'un facteur devrait être au moins égale au poids symbolique d'une variable (c.-à-d. une valeur de 1). D'ailleurs, selon Guttmann (1954), chaque facteur qui a une *valeur propre* supérieure à 1 devrait être retenu dans l'AFE (critère par défaut dans le logiciel SPSS). En fait, une valeur propre supérieure ou égale à 1 doit être considérée comme un point de référence et non comme une norme rigide en soi, car il est possible d'accepter, dans certains cas, une valeur propre légèrement inférieure à 1 (Thompson, 2004). Ces considérations chez le chercheur dépendront en partie des autres critères de l'AFE, mais aussi de l'assise théorique sous-jacente au phénomène étudié. Tabachnick et Fidell (2007) ajoutent que des valeurs propres sont plus grandes pour les quatre premiers facteurs et, après le sixième, ces valeurs sont souvent plus petites. Il y aurait donc en général un constat que la solution factorielle la plus probable se situe entre un et six facteurs. Il est aussi suggéré d'observer le test d'éboulis afin d'avoir une représentation visuelle des valeurs propres associées à chaque facteur.

Le *test d'éboulis* (*scree plot*) est un test graphique qui permet de déterminer le nombre de facteurs à retenir ou à extraire (Cattell, 1966). Il est approprié de tenir compte du nombre de facteurs en fonction de la *cassure* et de la valeur propre de chaque facteur. Le test d'éboulis commence en général à la suite du premier facteur (la valeur propre la plus élevée). Ensuite, il y a une pente considérable (ou cassure) des valeurs propres des facteurs subséquents, pour finir en une ligne horizontale. C'est pourquoi on décrira souvent le résultat graphique du test d'éboulis comme en forme de coude. Hair *et al.* (2006) suggèrent que le nombre de facteurs à retenir devrait se situer en amont de la cassure qui se dessine sur le test d'éboulis. Notons qu'il est parfois possible d'observer une ou plusieurs cassures sur le graphique du test d'éboulis, ce qui offre au chercheur plusieurs alternatives quant au nombre de facteurs à retenir. La difficulté pour déterminer le point de coupure le plus précis peut parfois amener le chercheur à surestimer le nombre de facteurs à extraire (Henson et Roberts, 2006). À ce sujet, Cattell (1966) mentionne qu'un des pièges est de retenir un facteur qui n'apporte que peu d'information ou de variance et qui, par conséquent, reste trivial pour l'explication du regroupement de certains observables. Il est à noter que la sélection du nombre de facteurs dépendra donc à la fois de la valeur propre du facteur, de la cassure (ou des cassures) apparente sur le test d'éboulis ainsi que du pourcentage de la *variance cumulée des facteurs*, laquelle devrait être égale ou supérieure à 50 %. Une variance cumulée des facteurs inférieure à 50 % indiquerait que les observables analysés dans l'AFE ne sont pas suffisants pour expliquer le phénomène à l'étude. Une analyse minutieuse de la variance ajoutée par chaque facteur à la variance totale de l'ensemble des facteurs permettra de préciser toujours un peu plus le choix du nombre de facteurs à retenir ou à extraire.

En plus de l'analyse des valeurs propres des facteurs de l'AFE et du test d'éboulis, il existe aussi l'*analyse parallèle*[4] (*parallel analysis*). Cette dernière est toutefois moins conventionnelle même si pertinente (car non incluse dans les logiciels comme SPSS)[5]. Elle fait référence à une simulation appelée Monte Carlo, dans laquelle les valeurs propres « attendues » sont obtenues en simulant des échantillons de façon aléatoire mettant en parallèle les valeurs propres de l'échantillon observé (Ledesma et Valero-Mora, 2007). La comparaison des valeurs propres (simulées et observées) permettra de tenir compte du nombre de facteurs à extraire.

4. Pour de plus amples informations, concernant cette analyse, consulter O'Connor, 2000.
5. Il existe des sites Internet à partir desquels il est possible de télécharger les applications requises pour effectuer les analyses parallèles.

1.4. Éléments à considérer après avoir fixé le nombre de facteurs à extraire de l'AFE initiale

À la suite de l'analyse fine des critères appartenant à la section précédente, notamment la valeur propre relative à chaque facteur, le test d'éboulis, la variance cumulée pour l'ensemble des facteurs et l'analyse parallèle (optionnelle), le chercheur pourra réaliser une nouvelle AFE (ou ACP) en *forçant* cette fois-ci le nombre de facteurs qu'il aura retenus. Tabachnick et Fidell (2001) recommandent que les chercheurs testent lorsque cela est approprié (p. ex., selon le nombre de cassure au test d'éboulis) un nombre différent de facteurs pour retenir la solution factorielle la plus satisfaisante. Dans cette section, le type de rotation des axes, les indices de saturation des observables à un facteur ainsi que l'appellation des facteurs pour représenter leurs contenus seront présentés.

La *rotation des facteurs* implique de mouvoir les axes factoriels dans l'espace, ce qui permet une meilleure localisation d'un ensemble d'observables sur un facteur afin que la nature du construit (ou du facteur) émergeant devienne plus facile à interpréter (Beavers *et al.*, 2013 ; O'Brien, 2007). La rotation est cependant impossible quand il n'y a qu'un seul facteur retenu. La rotation des facteurs est nécessaire afin d'obtenir une solution qui est ainsi caractérisée : *a*) chaque observable possède un niveau significatif de saturation à un seul facteur, et *b*) chaque facteur possède un niveau significatif de saturation pour seulement quelques indicateurs (Norušis et SPSS Inc., 2003 ; Thompson, 2004). Il existe deux types de rotation, orthogonale et oblique (Thompson, 2004). Une rotation orthogonale implique que les facteurs ne soient pas corrélés entre eux et, par conséquent, les axes de rotation sont placés dans l'espace à 90 degrés. Les méthodes de rotation orthogonales les plus connues sont Varimax, Quartimax et Equamarx[6]. Ces types de rotation font partie des logiciels SAS et SPSS (Kaiser, 1958 ; Thompson, 2004). Quant à la rotation oblique, elle fait référence à une corrélation entre les facteurs dont plusieurs méthodes existent telles que Direct Oblimin et Promax, pour celles qui sont les plus utilisées (Hendrickson et White, 1964 ; Thompson, 2004).

Lorsque comparée à la rotation orthogonale, celle qui est oblique contrôle pour la variance partagée entre les facteurs. Toutefois, Dassa (1996) ajoute que l'inconvénient de l'utilisation de la rotation oblique est qu'elle consiste en une forme de « cousinage » entre les facteurs qui émergeront de l'AFE, ce qui rend l'interprétation beaucoup plus ardue. Tabachnick et Fidell (1996, 2001) stipulent à ce propos que l'utilisation de différentes méthodes d'extraction, c'est-à-dire de types de rotation (ainsi que les méthodes de

6. Pour une revue de ces divers types de rotation, le lecteur est invité à consulter Browne, 2010.

réduction des données telles que l'ACP et la PAF vues plus haut), tend à donner des résultats similaires si le patron de corrélation entre les observables est clair et solide. À choisir entre deux rotations qui présentent des résultats similaires, la rotation orthogonale (p. ex., Varimax) restera probablement celle à retenir, sachant qui plus est que cette dernière permet de redistribuer la variance entre les facteurs afin de rééquilibrer le pourcentage de variance de l'AFE (Tabachnick et Fidell, 1996). De plus, parmi les types de rotation orthogonale, celle intitulée Varimax sera plus souvent retenue, car elle minimise la complexité des facteurs tout en optimisant la variance des saturations des observables à chaque facteur (Tabachnick et Fidell, 2007). Toutefois, Fabrigar *et al.* (1999) émettent quelques réserves à ce sujet, car dans le domaine de la psychologie ou des sciences humaines de façon plus générale, il y a des fondements théoriques qui supposent la présence de corrélations entres les concepts ou leurs sous-dimensions. Par conséquent, l'utilisation d'une rotation oblique peut représenter de façon plus réaliste et précise la façon dont les concepts (ou facteurs) sont reliés les uns aux autres et, du coup, la corrélation interfacteurs peut s'avérer mieux définie (Fabrigar *et al.*, 1999). Pour conclure, Henson et Roberts (2006) suggèrent que, quelle que soit la décision prise par le chercheur, le choix d'une rotation oblique ou orthogonale devra être clairement justifié. Dans cette veine, Henson et Roberts (2006) précisent que la rotation oblique pourrait tout d'abord être exécutée dans l'AFE, et dans la mesure où celle-ci indiquerait une corrélation interfacteurs qui n'est pas élevée ($r < 0,30$)[7], l'utilisation de la rotation orthogonale semblerait justifiée.

Les *indices de saturation* peuvent s'interpréter comme des coefficients de corrélation entre un facteur et un observable. Autrement dit, l'indice de saturation d'un observable à un facteur doit être lu comme sa contribution ou son poids à ce même facteur. On dira d'un observable qu'il sature au facteur X avec un coefficient d'une valeur Y (valeur qui peut varier de –1 à +1). L'indice de saturation d'un observable à un facteur, inférieur à –0,30 ou supérieur à 0,30, est considéré comme une contribution significative pour représenter ce facteur. Comme un indice de saturation de 0,30 en valeur absolue représente statistiquement une variance d'un peu moins de 10%, un coefficient plus élevé (p. ex., de l'ordre de 0,40) pourrait alors convenir selon l'étude et les objectifs que se sera fixés le chercheur. À ce propos, Thurstone (1938, cité dans Cudeck et O'Dell, 1994) rejetait des indices de saturation inférieurs à 0,20 et interprétait seulement les indices de saturation supérieurs à 0,40 comme significatifs. Tabachnick et Fidell (2007)

7. Il est à noter qu'il n'y a pas de normes établies pour décider qu'une corrélation est jugée assez élevée. À l'instar de ce qui se fait avec les indices de saturation, O'Brien (2007) suggère qu'une corrélation de 0,30 est un seuil minimal pour qu'un coefficient de corrélation soit considéré comme suffisamment élevé.

concluent que l'interprétation de la taille des indices de saturation dépend avant tout des objectifs du chercheur et du contexte de l'étude. Outre l'ensemble de ces subtilités, Comrey et Lee (1992) suggèrent de qualifier les indices de saturation aux facteurs en considérant les seuils ci-dessous :

- indice de saturation de l'ordre de 0,71 (50 % de variance) = seuil excellent ;
- indice de saturation de l'ordre de 0,63 (40 % de variance) = seuil très bon ;
- indice de saturation de l'ordre de 0,55 (30 % de variance) = bon seuil ;
- indice de saturation de l'ordre de 0,45 (20 % de variance) = seuil acceptable ;
- indice de saturation de l'ordre de 0,32 (10 % de variance) = seuil faible.

Par ailleurs, rappelons que l'AFE peut être utilisée pour éliminer des observables qui semblent moins pertinents pour l'explication du phénomène étudié (application de la loi de la parcimonie). Toutefois, prendre la décision de conserver ou non un observable dans une AFE peut avoir des répercussions sur l'ensemble de la solution factorielle, et il devient donc important d'évaluer les tenants et aboutissants de cette décision. La décision de retrancher ou non un observable d'une AFE se prendra, notamment, après la considération des critères suivants (non exhaustifs) : la pertinence clinique de l'observable dans la compréhension d'un phénomène et la contribution significative de l'observable à un seul facteur. Une double saturation d'un observable à deux facteurs doit toujours être analysée afin de bien l'interpréter. À la suite de cette analyse plus fine, le chercheur prendra la décision d'assigner l'observable à un facteur plutôt qu'à un autre ou encore de retrancher l'observable de l'analyse factorielle, car il s'avère bancal ou diffus sur le plan théorique ou clinique.

Comme nous avons pu le constater à plusieurs reprises, le contenu d'un facteur est étroitement lié aux observables qui le composent. Et, par conséquent, l'*intitulé des facteurs* devrait refléter la contribution de l'ensemble des observables pour bien définir ces premiers. D'ailleurs, il est important de bien nommer les facteurs avec des intitulés concis les plus précis possibles afin de mener ultérieurement des comparaisons de résultats. L'intitulé du facteur devrait faire référence non seulement au concept à évaluer dans le questionnaire (ou outil de mesure) en reflétant sa dimension conceptuelle plus générale, mais aussi au contenu des observables dont la contribution

est la plus élevée (O'Brien, 2007). Par ailleurs, il est aussi suggéré d'avoir au moins trois observables pour représenter un facteur ou une dimension conceptuelle (Lent, Brown et Gore, 1997 ; Quintana et Maxwell, 1999). Ainsi, le facteur ne pourra être que plus stable et solide conceptuellement et statistiquement parlant (Costello et Osborne, 2005).

Pour conclure, l'AFE est devenue très populaire auprès de la communauté scientifique. Russell (2002) mentionne que ce type d'analyse est utilisé dans plus d'un quart des articles publiés dans certaines revues scientifiques. Cependant, Pedhazur et Pedhazur Schmelkin (1991) nous mettent en garde concernant l'utilisation de l'AFE, car cette dernière est une forêt dans laquelle il est facile de se perdre en tout temps. Il est donc suggéré, lors de l'utilisation d'une AFE, d'avoir un cadre théorique sur lequel s'appuyer pour analyser les observables. Une AFE n'est donc jamais complètement dénuée d'une assise théorique et, par conséquent, elle n'est jamais tout à fait exploratoire (Thompson, 2004). De plus, Henson et Roberts (2006) mentionnent que la limite la plus fréquemment citée vis-à-vis de l'AFE est le niveau de subjectivité du chercheur dans les décisions méthodologiques, qui peuvent influencer la qualité des résultats de cette analyse. En bref, l'utilisation et l'interprétation des résultats d'une AFE exigent chez le chercheur de réaliser un travail systématique et rigoureux tout en faisant preuve d'un sens commun aiguisé, le tout étayé par un fondement empirique ou théorique solide auquel il peut se référer (voir le tableau 22.1 pour un résumé des éléments à considérer). En ce sens, l'AFE et l'AFC ont des points en commun.

Tableau 22.1.
Aspects essentiels à considérer pour l'AFE

Aspects à considérer pour l'AFE	Informations particulières
Avant l'utilisation d'une AFE	
▪ Existence d'une assise théorique ou empirique	Un travail antérieur a été réalisé sur le concept à évaluer ou le phénomène à l'étude.
▪ Taille de l'échantillon	5 à 10 sujets pour chaque variable
	N > 150 (ou justification quand l'échantillon est plus petit)
▪ Analyses préliminaires des observables (ou données)	Vérification de la multicolinéarité, de la non-normalité des observables, des données aberrantes et manquantes
▪ Coefficients de corrélation entre les observables	Les coefficients devraient être égaux ou supérieurs à 0,30, mais inférieurs à 0,70.

(*suite*)

Tableau 22.1.
Aspects essentiels à considérer pour l'AFE (*suite*)

Aspects à considérer pour l'AFE	**Informations particulières**
Réalisation de l'AFE initiale	
▪ Choix de la technique utilisée	Analyse par composante principale (ACP) ou analyse factorielle exploratoire (AFE)
▪ Bartlett's Test of Sphericity	Valeur du $p < 0{,}05$
▪ Kaiser-Meyer-Olkin measure of sampling adequacy (KMO)	Coefficient > 0,60
Évaluation et décision du nombre de facteurs à extraire de l'AFE initiale	
▪ Valeur propre (*eigenvalue*)	> 1 (une valeur inférieure, mais proche de 1 est possible)
▪ Variance totale cumulée	≥ 50 %
▪ Test d'éboulis (*scree test*)	Observation des « cassures » et des valeurs propres
▪ Analyse parallèle (*parallel analysis*)	Comparaison des résultats avec des échantillons aléatoires (simulation) ; analyse optionnelle accessible sur Internet
Réalisation d'une nouvelle AFE en spécifiant le nombre de facteurs à extraire	
▪ Type de rotation des axes factoriels	Orthogonal (p. ex. Varimax) ou oblique (p. ex., Oblimin)
▪ Indice de saturation d'un observable à un facteur (*loading*)	≥ 0,30 ou –0,30 ≤ Cependant, certains auteurs suggèrent des indices de saturation égaux ou supérieurs à 0,40.
▪ Décision d'éliminer ou pas des observables de l'AFE	Elle dépendra des objectifs du chercheur, de la pertinence des observables sur le plan clinique ou théorique, ainsi que des indices de saturation (faible ou double saturation). Si certains items sont éliminés, alors le chercheur devra effectuer à nouveau l'AFE en retirant ces observables de l'analyse.

2. INTRODUCTION À L'ANALYSE FACTORIELLE CONFIRMATOIRE

L'analyse factorielle confirmatoire (AFC) est un cas particulier de la modélisation par équations structurelles (Mueller, 1996). Lorsque les chercheurs utilisent la modélisation par équations structurelles (voir le chapitre 21 de cet ouvrage), il n'est pas rare qu'ils préconisent en premier lieu l'AFC pour tester le modèle de mesure sous-jacent à un concept ou un construit (*measurement model*). En effet, cette étape préliminaire permet d'assurer la validité des concepts, pour ensuite les intégrer et les articuler dans un modèle théorique plus complexe (Asparouhov et Muthén, 2009; Babyak et Green, 2010; Jackson, Gillaspy et Purc-Stephenson, 2009). Dans cet ordre d'idées, une AFE peut aussi être préliminaire à l'AFC au sens où l'AFE est réalisée

dans une première étude (ou avec un premier échantillon) et fournit ainsi les bases nécessaires pour spécifier un modèle AFC dans une étude subséquente (ou avec un second échantillon) (Fabrigar *et al.*, 1999; Jackson *et al.*, 2009). Ainsi, de fil en aiguille, les chercheurs auront accès à un volume de connaissances plus substantiel sur lequel ils pourront étayer leurs nouveaux objectifs de recherche.

L'AFC se distingue de l'AFE (vue plus haut), car cette première exige du chercheur d'émettre des attentes précises en matière de résultats, notamment le nombre de facteurs à retenir (s'ils sont corrélés ou non) et le nombre d'observables qui saturent au(x) facteur(s). Autrement dit, dans le cadre d'une AFC, tous les paramètres implicites du modèle de mesure doivent être estimés. Le chercheur peut établir certaines contraintes ou fixer certains paramètres à estimer, et peut en libérer d'autres au sens où une estimation est requise (voir le chapitre 21 de cet ouvrage). Les chercheurs sans théorie ou sans cadre conceptuel précis ne peuvent utiliser l'AFC pour tester leurs données. Lorsque le chercheur a précisé ses attentes vis-à-vis du modèle de mesure qu'il souhaite tester, il peut alors envisager de conduire une AFC. Enfin, DiStefano et Hess (2005) recommandent lors de l'utilisation des AFC de tester des modèles de mesure alternatifs, ce qui permet d'éviter certains *biais de confirmation*, où un modèle de mesure pourrait être bien ajusté aux données testées sans la vérification d'autres modèles de mesure aussi pertinents ou équivalents, voire plus solides.

L'évaluation de l'ajustement du modèle aux données empiriques est réalisée en fonction de plusieurs indices d'ajustement[8] tels que le chi-carré, le ratio chi-carré/degré de liberté, le RMSEA (Root Mean Square Error of Approximation), le NNFI (Non-Normed Fit Index), le CFI (Comparative Fit Index), l'IFI (Incremental Fit Index) (Bentler, 1995; Jöreskog et Sörbom, 1993; Hooper, Coughlan et Mullen, 2008; Mueller, 1996) pour n'en nommer que quelques-uns. Quant au seuil à respecter pour ces indices, Schermelleh-Engel, Moosbrugger et Müller (2003) proposent un tableau des coefficients avec des bornes particulières, où $0{,}95 \leq \text{NFI} \leq 1{,}00$ représente un bon ajustement alors que la borne $0{,}90 \leq \text{NFI} < 0{,}95$ fait référence à un ajustement acceptable. Bentler (1995) propose de considérer deux tests additionnels pour améliorer le modèle hypothétique: le multiplicateur de Lagrange et le test de Wald.

Le test de Wald permet d'évaluer l'effet du *rejet* d'un paramètre à estimer (p. ex., la corrélation entre deux variables ou deux facteurs ou deux erreurs de mesure) sur le modèle hypothétique. Le multiplicateur de

8. Ces indices d'ajustement sont présentés et définis dans le chapitre 21 de cet ouvrage, qui porte sur la modélisation par équations structurelles.

Lagrange consiste à évaluer l'effet de l'*ajout* d'un paramètre à estimer sur le modèle hypothétique. Ces deux tests additionnels sont donc utiles pour comparer des modèles nichés au modèle hypothétique en vue de repérer un modèle mal spécifié. Autrement dit, d'un mode confirmatoire, il est possible d'explorer de nouveaux chemins empiriques ou encore, d'une démarche hypothético-déductive, il est possible de passer à une démarche heuristique (Bacher, 1987). Il est à noter ici que ce ne sont pas toutes les écoles de pensée qui adhèrent à cette démarche heuristique. Cependant, elle offre aux chercheurs une flexibilité en vue d'ouvrir sur des voies de recherche innovatrices ou tout simplement pour raffiner le modèle de mesure testé. Enfin, la méthode d'estimation la plus répandue pour tester des modèles à l'aide d'une AFC est la méthode de maximum de vraisemblance (Jöreskog et Sörbom, 1993) ou la méthode de maximum de vraisemblance dite robuste (Bentler, 1995). Cette dernière est utilisée pour pallier la violation de l'hypothèse de la normalité multivariée. Hu et Bentler (1999) recommandent pour des échantillons inférieurs à 250 individus d'utiliser la méthode de maximum de vraisemblance dite robuste pour contrôler la violation de la normalité des variables.

Pour ce qui est de la taille de l'échantillon requise pour ce type d'analyse, plusieurs auteurs stipulent qu'un échantillon de 200 individus et plus semble adéquat pour respecter les exigences des indices d'ajustements susmentionnés (Bollen, 1989 ; Chou et Bentler, 1995). Toutefois, un échantillon plus conséquent est plus souvent recommandé (Quintana et Maxwell, 1999). En se référant à de nombreux auteurs, Schermelleh-Engel *et al.* (2003) stipulent que la taille d'échantillon minimale peut osciller entre 100 et 300 individus en fonction de la distribution normale des données, de la spécification du modèle de mesure et du nombre d'observables par facteur. Ils ajoutent que le nombre d'observables par facteur (plus de six observables) pourrait compenser pour des échantillons de petite taille et, à l'inverse, des échantillons de plus grande taille peuvent compenser par un faible nombre d'observables par facteur (trois à quatre observables). À l'instar de ce qui est fait dans l'AFE, DiStefano et Hess (2005) mentionnent que l'évaluation d'un concept à l'aide de un ou deux observables peut causer des problèmes d'estimation et d'interprétation du concept à l'étude. Il est donc fortement recommandé d'utiliser trois, voire quatre observables et plus par facteur lors de l'utilisation d'une AFC, notamment pour les modèles de mesure (Fabrigar *et al.*, 1999).

Il existe plusieurs logiciels pour effectuer une AFC, comme EQS (Bentler, 1995), LISREL (Jöreskog et Sörbom, 1996), Mplus (Muthén et Muthén, 1998) et AMOS (Arbuckle et Wothke, 1999). En plus des recommandations apportées dans le chapitre sur la modélisation par équations

structurelles (voir le chapitre 21 de cet ouvrage), il existe dans la littérature spécialisée des listes de vérification (*checklists*) pour conduire des AFC (DiStefano et Hess, 2005 ; Jackson *et al.*, 2009). Nous résumons dans le tableau 22.2 les principaux éléments relevés par Jackson *et al*. (2009).

Tableau 22.2.
Guide pour les AFC

Formulation théorique et collecte des données

- La justification théorique ou empirique du modèle (ou modèles) à tester ;
- Le nombre de modèles à tester (p. ex., facteurs corrélés ou non) ;
- La spécification des modèles à tester (les relations entre les observables et les concepts sont explicites) ;
- La représentation graphique des modèles à tester, notamment lorsqu'ils sont complexes ;
- La description des caractéristiques de l'échantillon (p. ex., taille de l'échantillon) ;
- L'identification des modèles alternatifs sur le plan théorique.

La préparation des données

- L'identification de la non-normalité des données multivariées et des données aberrantes ;
- L'analyse des données manquantes en vue de les traiter ;
- La description de l'échelle utilisée pour les observables (p. ex., ordinale, intervalle) ;
- La description de la transformation des observables au besoin (p. ex., *parcelling*).

Les décisions prises pour conduire l'analyse

- La description de la matrice analysée (covariance ou corrélation) ;
- La matrice est incluse ou disponible sur requête ;
- La justification de la procédure d'estimation (p. ex., maximum de vraisemblance) ;
- La description de l'échelle pour le concept évalué (ou variable latente) ;
- La description sommaire du logiciel utilisé (p. ex., LISREL, EQS, AMOS).

L'évaluation du ou des modèles

- L'inclusion de plusieurs indices d'ajustement (p. ex., chi-carré, RMSEA, CFI).

Source : Le contenu de ce tableau est tiré de Jackson, Gillaspy et Purc-Stephenson, 2009.

3. APPLICATION DES ANALYSES FACTORIELLES EXPLORATOIRES ET CONFIRMATOIRES

Dans cette partie, il s'agira d'illustrer les deux types d'analyses factorielles, exploratoire et confirmatoire. Pour ce faire, il a été choisi de réaliser les AFE et AFC dans le cadre de la validation d'un questionnaire intitulé l'Échelle d'estime de soi en tant que travailleur – ESST (Corbière *et al.*, 2009), et ce, auprès de personnes aux prises avec un trouble mental grave inscrites

dans des programmes canadiens de soutien à l'emploi. Pour chacune de ces analyses, les éléments qui ont été présentés dans la première partie de ce chapitre seront repris de façon appliquée, d'abord avec l'AFE puis avec l'AFC. L'objectif de cette illustration est aussi de montrer au lecteur la pertinence de combiner ces deux analyses pour la validation d'un questionnaire. Les logiciels SPSS et EQS ont été utilisés pour réaliser respectivement l'AFE et l'AFC.

Les données ou observables du projet de recherche, dont l'objectif était de mieux comprendre la réinsertion au travail de personnes aux prises avec un trouble mental grave, ont été recueillies auprès de 606 personnes inscrites dans des programmes canadiens de soutien à l'emploi (Corbière *et al.*, 2006). On compte parmi elles 384 personnes provenant des provinces de la Colombie-Britannique et de l'Ontario et 222 personnes de la province de Québec. L'échantillon anglophone ($n = 384$) sera utilisé aux fins de l'AFE, et l'échantillon francophone ($n = 222$) pour l'AFC (pour plus de détails sur la validation d'un outil, voir le chapitre 24 de cet ouvrage). L'ESST a été validé dans une autre étude auprès d'un échantillon anglophone de même nature (Corbière *et al.*, 2009) et il sera question de répliquer ce travail de validation avec des données provenant d'un nouveau projet de recherche.

L'ESST est composé de dix items (échelle de réponse de 1 = tout à fait en désaccord à 4 = tout à fait d'accord), dont cinq énoncés sont formulés à la positive et cinq autres à la négative. L'ESST a été conçu à partir de l'Échelle d'estime de soi de Rosenberg (1965), notamment en rajoutant dans la consigne et dans les énoncés de l'outil la notion de travail[9]. Cette adaptation a été réalisée parce que l'Échelle d'estime de soi de Rosenberg était trop générale pour évaluer la personne dans ce contexte de transition particulier et, par conséquent, pour rendre compte d'éventuels changements chez la personne avec un trouble mental grave avant et après l'obtention d'un emploi sur le marché du travail ordinaire (Corbière *et al.*, 2009). L'estime de soi de Rosenberg comme l'estime de soi évaluée par d'autres questionnaires (Corbière *et al.*, 2009 ; Lecomte, Corbière et Laisné, 2006 ; Rosenberg, 1965) ont montré que la solution factorielle présentait soit une seule et unique dimension conceptuelle (Estime de soi générale – 10 énoncés), soit une double dimension, positive (cinq énoncés formulés à la positive) et négative (cinq énoncés formulés à la négative), ou encore une double dimension intitulée *estime de soi en tant que travailleur de nature individuelle* (évaluation de son estime par rapport à soi – sept énoncés) et une autre intitulée *estime de soi en tant que travailleur de nature sociale* (évaluation de son estime par

9. Par exemple, l'énoncé 1 : *Je pense que je suis une personne de valeur, au moins égale à n'importe quelle autre personne* de l'Échelle d'estime de soi de Rosenberg a été adapté au contexte du travail : *En tant que travailleur, je pense que je suis une personne de valeur, au moins égale à n'importe quel autre travailleur*.

rapport à autrui – trois énoncés). En bref, l'estime de soi adaptée au contexte du travail est un concept pertinent pour l'évaluation des personnes aux prises avec un trouble mental en processus de recherche d'emploi et peut présenter trois solutions factorielles.

Avant de réaliser les analyses factorielles (exploratoire et confirmatoire), la banque de données a été vérifiée pour cerner d'éventuels problèmes de collecte de données (p. ex., données aberrantes ou manquantes). De ces observations préliminaires, seule ressortait la présence de moins de 1 % de données manquantes pour l'ensemble de l'échantillon (n = 606), ce qui n'indiquait aucun patron de réponse pour les données manquantes. Par ce très faible pourcentage, les données manquantes ont été remplacées par la moyenne du groupe (anglophone ou francophone) en considérant l'observable (ou item) particulier.

3.1. Illustration de l'AFE

Cette illustration suivra le découpage présenté en première partie du chapitre : avant l'utilisation de l'AFE, lors de l'utilisation d'une AFE initiale, au moment de fixer le nombre de facteurs à extraire de l'AFE initiale, et après avoir fixé le nombre de facteurs à extraire, une nouvelle AFE est réalisée.

3.1.1. Avant l'utilisation de l'AFE

Comme il est fortement recommandé, l'assise théorique et les résultats empiriques concernant le modèle de mesure à évaluer doivent être considérés, ce qui a été présenté un peu plus haut. Comme une AFE n'est jamais dénuée d'une assise théorique ou empirique, l'hypothèse relative à la solution factorielle de l'AFE était que l'une des trois solutions factorielles présentées ci-dessus pouvait émerger de l'ESST (voir ci-dessus). Toutefois, il est à noter que comme l'analyse factorielle est de type exploratoire, les résultats de l'AFE peuvent soit soutenir l'une ou l'autre de ces trois solutions factorielles, soit encore dégager une quatrième solution factorielle, propre à l'échantillon à l'étude.

La taille de l'échantillon est de 384 personnes anglophones avec un trouble mental grave qui ont répondu à l'ESST (version anglaise). Cette taille respecte les exigences du ratio observable/nombre d'individus de 1 pour 10 et du minimum requis de 150 individus. La matrice de corrélation de Pearson a été observée et il ressort que la plupart des observables (c.-à-d. les

énoncés de l'ESST) présentent une corrélation significative (au seuil $p < 0,05$) dont le coefficient le plus bas est de 0,26 et celui qui est le plus élevé est de 0,59, ce qui respecte les critères requis pour exécuter une AFE[10].

3.1.2. Lors de l'utilisation d'une AFE initiale

Pour ce qui est du choix de la technique, dans cette illustration, l'ACP (c.-à-d. l'analyse par composantes principales) comme l'AFE (c.-à-d. la factorisation des axes principaux) seront retenues afin d'observer les résultats et de les comparer. Par ailleurs, les résultats de la sortie de résultats 1 (tableau 22.3) indiquent que les valeurs du KMO et du test de Bartlett sont significatives, puisque le premier indice est supérieur à 0,90 (coefficient excellent ; Kaiser, 1974) et le second est significatif pour une valeur de $p < 0,001$. Ces résultats signifient donc que les données se prêtent bien à l'AFE. Il est à noter que ces valeurs sont identiques quelle que soit la technique utilisée (ACP ou AFE).

Tableau 22.3.
Sortie des résultats 1

Valeur du KMO et du test de Bartlett		
Mesure de précision de l'échantillonnage de Kaiser-Meyer-Olkin		0,917
Test de sphéricité de Bartlett	chi-carré approximé	1576,864
	ddl	45
	signification de Bartlett	0,000

3.1.3. Au moment de fixer le nombre de facteurs à extraire de l'AFE initiale

Pour se prêter à l'exercice de la comparaison des résultats de l'ACP à ceux de l'AFE, nous avons réalisé ces deux analyses (Pett, Lackey et Sullivan, 2003). Dans le tableau 22.4, d'une part, il est possible d'observer que seulement deux facteurs présentent des valeurs propres supérieures à 1 (c.-à-d. 4,967 et 1,089), ce qui indique que deux facteurs pourraient être retenus dans la solution factorielle finale (Guttmann, 1954). D'autre part, lorsque le pourcentage de variance cumulée des deux facteurs de l'ACP (60,558 %) est comparé à celui de l'AFE (50,738 %), on note que dans le premier cas, il est

10. Avant le calcul des coefficients de corrélation, les scores des énoncés formulés à la négative ont été inversés, comme suggéré par Rosenberg (1965) où 1 = 4, 2 = 3, 3 = 2 et 4 = 1. Ceci explique que l'on obtienne exclusivement des coefficients de corrélation positifs.

plus élevé (10% de plus). Ce dernier résultat reflète le fait que la variance totale est considérée dans l'ACP (c.-à-d. la variance totale des observables) alors que seule la variance commune l'est dans l'AFE (c.-à-d. la variance commune partagée par les observables). Dans les deux types d'analyse, le pourcentage de variance cumulée est supérieur à 50% et respecte ainsi le seuil minimal requis.

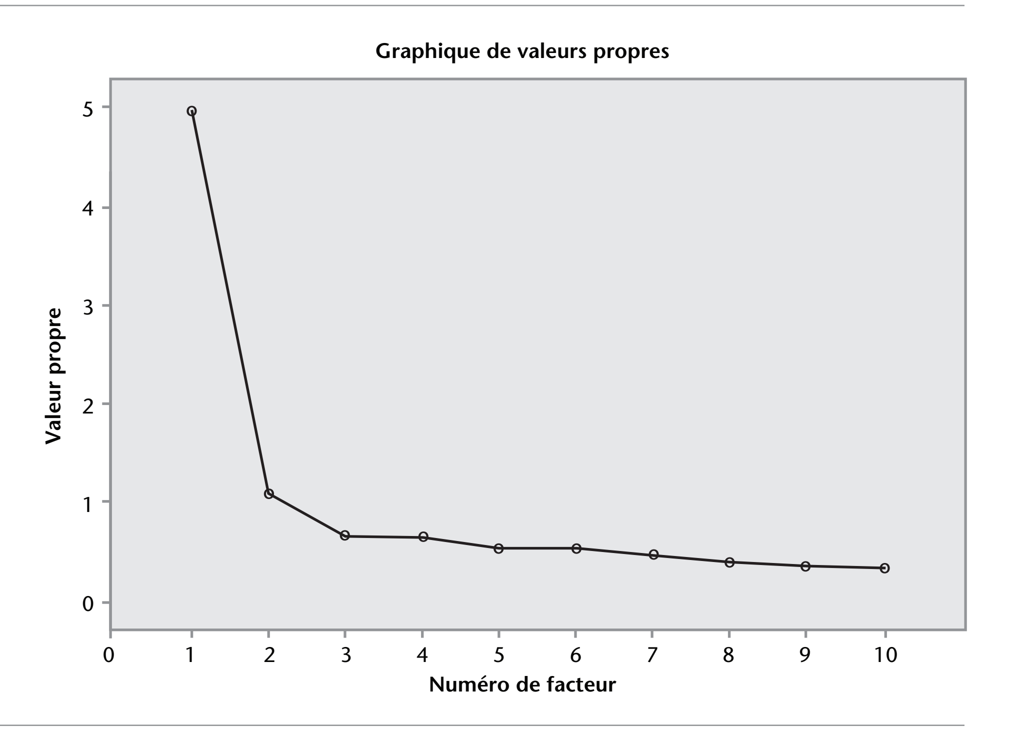
Tableau 22.4.
Résultats selon la méthode ACP et celle de la factorisation des axes principaux (ou AFE)

Valeurs propres initiales (méthodes ACP et factorisation des axes principaux)				Extraction – Sommes des carrés des facteurs retenus					
				Méthode ACP			Méthode de factorisation des axes principaux		
Composante	Total	% de la variance	% cumulés	Total	% de la variance	% cumulés	Total	% de la variance	% cumulés
1	4,967	49,667	49,667	4,967	49,667	49,667	4,486	44,858	44,858
2	1,089	10,891	60,558	1,089	10,891	60,558	0,588	5,880	50,738
3	0,655	6,546	67,104						
4	0,637	6,368	73,472						
5	0,534	5,341	78,813						
6	0,521	5,207	84,021						
7	0,472	4,722	88,743						
8	0,398	3,983	92,726						
9	0,381	3,806	96,531						
10	0,347	3,469	100,000						

On observe dans le test d'éboulis (sortie de résultats 2, figure 22.1) qu'une *cassure* est présente après le second facteur (forme de coude), et que les autres facteurs suivent une ligne horizontale. On peut observer aussi une *cassure* entre le premier et le second facteur, ce qui signifie qu'un seul facteur pourrait être retenu dans l'analyse factorielle exploratoire. Toutefois, la solution à un seul facteur présenterait un pourcentage de variance insuffisant pour être adoptée. Par conséquent, la solution factorielle à deux facteurs semble la plus appropriée pour l'échantillon à l'étude. Enfin, comme l'analyse parallèle est optionnelle et exige l'utilisation d'un logiciel qui n'est

pas toujours accessible à tous, elle n'a pas été réalisée dans cette illustration. La prochaine étape de l'AFE prendra donc en compte ces résultats en considérant l'analyse factorielle exploratoire de type factorisation par axes principaux[11].

Figure 22.1.
Sortie des résultats 2 – Test d'éboulis



3.1.4. Réalisation d'une nouvelle AFE en spécifiant le nombre de facteurs à extraire

Dans cette dernière étape, l'objectif est de *forcer* l'AFE à deux facteurs en précisant le type de rotation des axes qui sera utilisé[12], soit orthogonale (c.-à-d. aucune corrélation entre les facteurs n'est attendue), soit oblique (c.-à-d. une corrélation entre les facteurs est attendue). Dans le cadre de

11. L'ACP aurait pu être choisie si le chercheur visait l'évaluation de la variance totale des observables. Dans cette illustration, nous avons choisi la méthode par factorisation des axes principaux, car elle représente le standard à suivre pour l'analyse factorielle exploratoire.
12. En l'occurrence, comme il y a plus d'un facteur, il est possible de mouvoir les axes factoriels.

cette illustration, les deux types de rotation ont été exécutés (tableau 22.5), par rotation Varimax (orthogonale) et par rotation Oblimin (oblique). Par ailleurs, une valeur égale ou supérieure à 0,40 pour les indices de saturation est retenue comme seuil critique dans cette étude, car le postulat de base est que les énoncés de l'ESST devraient contribuer significativement à leur facteur respectif (>10 %). Lorsque l'on compare les résultats entre les deux types de rotations (oblique et orthogonale), les résultats sont sensiblement similaires : les indices de saturation des énoncés 3, 5, 7, 9 et 10 sont significatifs (> 0,40) sur le premier facteur alors que les énoncés 1, 2, 4, 6 et 8 sont significatifs (> 0,40) sur le second facteur[13]. Il est à noter cependant que pour la solution à rotation orthogonale (Varimax), les énoncés 6 et 8 présentent une double saturation significative aux deux facteurs, ce qui indique que ces deux items présentent aussi un poids significatif aux deux facteurs de l'ESST, quoique la saturation la plus importante (en valeur absolue) de ces deux items reste plus importante pour le second facteur. Ces derniers résultats seront repris lors de l'illustration de l'AFC. Enfin, comme les deux facteurs présentent une corrélation significative (r = 0,687) (sortie des résultats 3, tableau 22.6) et que la solution par rotation Varimax semble moins intéressante sur le plan des indices de saturation du second facteur (deux items qui présentent une double saturation), la solution factorielle finale retenue est celle pour laquelle la rotation oblique (Oblimin) a été utilisée, ce qui indique une solution factorielle à deux dimensions intercorrélées, c'est-à-dire que les énoncés formulés à la positive regroupés sont corrélés à ceux formulés à la négative regroupés sous l'autre facteur[14].

13. Dans le tableau 22.5, la matrice des types (*pattern matrix*) pour la rotation Oblimin a été choisie, car elle est plus facile à comparer aux résultats obtenus lors de l'utilisation de la rotation Varimax. Mais la matrice structure (*structure matrix*) pourrait être également considérée dans une analyse plus fine de la rotation Oblimin. Pour de plus amples informations concernant ces deux types de matrice issus de la rotation Oblimin, le lecteur est invité à consulter les écrits de Tabachnik et Fidell (1996, 2001, 2007).
14. La corrélation entre les facteurs est positive, car le score des énoncés a été inversé avant l'utilisation des AFE.

Tableau 22.5.
Résultats de l'AFE avec rotation orthogonale et oblique

	Matrice des types[a]		Matrice factorielle après extraction[b]	
	Facteur		Facteur	
	1	2	1	2
Item 9	0,76		0,72	
Item 5	0,72		0,65	
Item 10	0,69		0,66	
Item 7	0,63		0,56	
Item 3	0,55		0,58	
Item 1		0,80		0,72
Item 2		0,75		0,69
Item 4		0,63		0,60
Item 6		0,49	0,44	0,56
Item 8		0,45	0,50	0,55

[a] Méthode d'extraction : factorisation en axes principaux. Méthode de rotation : Oblimin avec normalisation de Kaiser. La rotation a convergé en sept itérations.
[b] Méthode d'extraction : factorisation en axes principaux. Méthode de rotation : Varimax avec normalisation de Kaiser. La rotation a convergé en trois itérations.

Tableau 22.6.
Sortie des résultats 3 – Rotation Oblimin avec l'utilisation de l'AFE

Matrice de corrélation factorielle		
Facteur	1	2
1	1,000	0,687
2	0,687	1,000

3.2. Illustration de l'AFC

Dans le cadre de cette illustration, trois modèles de mesure ont été testés : 1) une seule et unique dimension conceptuelle (dix énoncés regroupés) ; 2) une double dimension, positive (cinq énoncés : esst1, esst2, esst4, esst6, esst8) et négative (cinq énoncés : esst3, esst5, esst7, ess9, esst10), toutes deux corrélées ; et 3) une double dimension intitulée *estime de soi en tant que travailleur de nature individuelle* (sept énoncés : esst3, esst5, esst6 à esst10) et

une autre intitulée *estime de soi en tant que travailleur de nature sociale* (trois énoncés : esst1, esst2 et esst4), toutes deux corrélées. Le tableau 22.7 montre que les deux modèles avec une double dimension s'ajustent mieux aux données empiriques que le modèle à dimension unique. De plus, même si les indices d'ajustement sont acceptables pour ces deux derniers modèles, il apparaît que le modèle à double dimension, individuelle et sociale, est plus solide, car les indices sont dans la borne plus satisfaisante, tant pour le ratio chi-carré/degrés de liberté (< 2) que pour les indices NNFI, CFI et IFI, tous supérieurs à 0,95, et le RMSEA est égal à 0,05. Qui plus est, l'indice AIC a une valeur plus basse pour ce dernier modèle (–15,20), ce qui indique qu'il est plus parcimonieux comparativement aux deux autres modèles. Lorsque les coefficients standardisés (équivalents aux indices de saturation) des énoncés à leur facteur respectif sont observés, on note que la quasi-totalité des énoncés sont supérieurs à 0,62 (sortie de résultats 4, figure 22.2) et qu'un seul énoncé présente un coefficient plus faible, mais significatif pour le poids qu'il représente au facteur, soit l'énoncé 7 (0,30). Enfin, la corrélation entre les deux facteurs est également significative avec un coefficient de 0,79[15]. L'ensemble de ces derniers coefficients est repris dans la figure 22.3. Ce type de figure (ainsi que le tableau 22.7) est d'ailleurs fortement recommandé lorsque le chercheur souhaite publier les résultats de l'AFC dans une revue scientifique.

Tableau 22.7.
Résultats des analyses factorielles confirmatoires pour les divers modèles testés

Modèles	Indices d'ajustement								
	ddl	**χ^2**	**Valeur p**	**χ^2/ddl**	**NNFI**	**CFI robuste**	**IFI**	**AIC**	**RMSEA**
M1. Une seule dimension	35	94,14	0,0001	2.69	0,88	0,91	0,91	24,14	0,09
M2. Deux dimensions positive et négative corrélées	34	73,35	0,001	2,16	0,92	0,94	0,94	5,35	0,07
M3. Deux dimensions individuelle et sociale corrélées	34	52,80	0,02	1,55	0,96	0,97	0,97	-15,20	0,05

Légende : NNFI = Non-Normed Fit Index ; CFI = Comparative Fit Index ; IFI = Bollen Incremental Fit Index ; AIC = Akaike Information Criterion ; RMSEA = Root Mean Square Error of Approximation.

15. Il est possible de vérifier si ces coefficients sont significatifs en consultant d'autres sections de la sortie de résultats de l'analyse, dans lesquelles les paramètres sont estimés (voir aussi le chapitre 21 de cet ouvrage).

Figure 22.2.
Sortie des résultats 4 EQS – Double dimension, individuelle (sept énoncés) et sociale (trois énoncés)

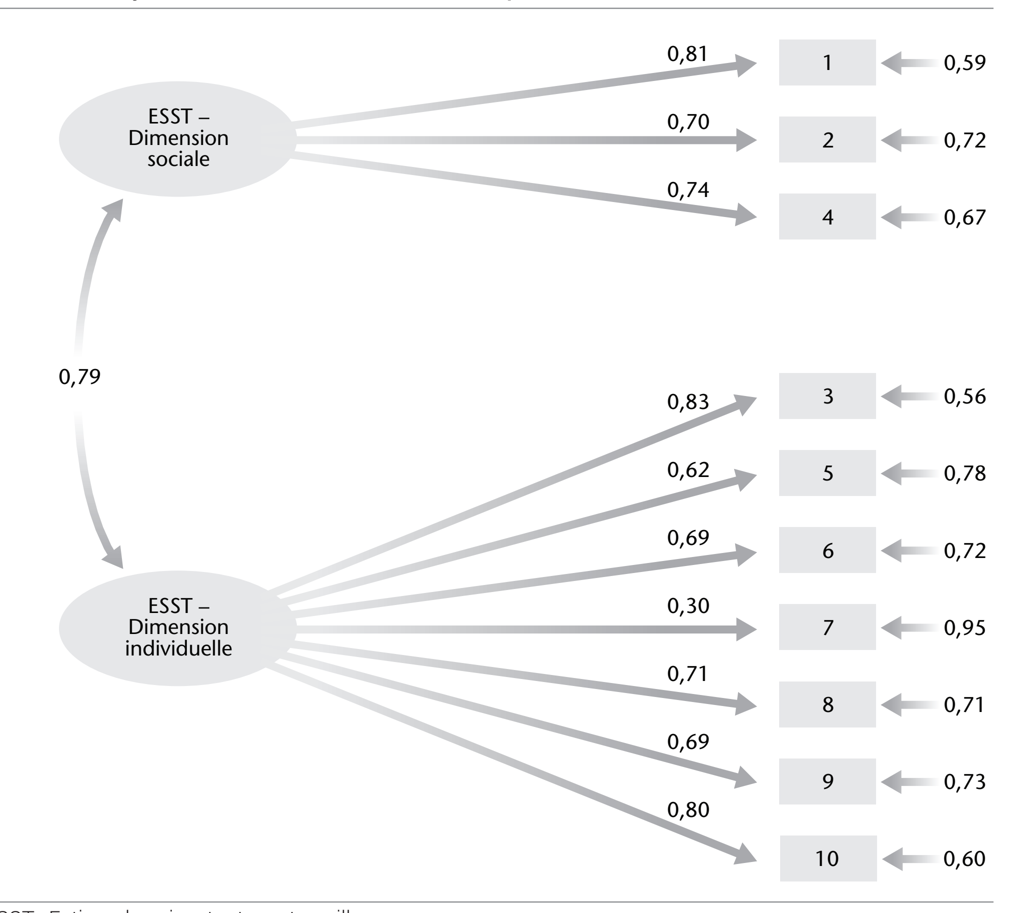
```
STANDARDIZED SOLUTION : R-SQUARED
ESTT1 = V1 =.808 F1 +.590 E1.652
ESTT2 = V2 =.696*F1 +.718 E2.485
ESTT4 = V3 =.744*F1 +.668 E3.553
ESTT3 = V4 =.831 F2 +.556 E4.691
ESTT5 = V5 =.622*F2 +.783 E5.387
ESTT6 = V6 =.689*F2 +.724 E6.475
ESTT7 = V7 =.302*F2 +.953 E7.091
ESTT8 = V8 =.706*F2 +.709 E8.498
ESTT9 = V9 =.688*F2 +.725 E9.474
ESTT10 = V10 =.800*F2 +.600 E10.641
CORRELATIONS AMONG INDEPENDENT VARIABLES
________________
          V                F
     ______        __________
          I     F2 – F2.786*I
          I          F1 – F1 I
          I                  I
```

Eu égard à l'ensemble des résultats, soit ceux de l'AFE et de l'AFC, on comprend que l'AFE permet de faire ressortir une solution factorielle à double dimension, positive et négative, alors que l'AFC apporte plus de nuances quant à la solution factorielle la plus solide des trois modèles testés. En effet, les deux modèles qui incluent une double dimension sont acceptables en matière d'indice d'ajustement, mais celui qui comprend les dimensions individuelle et sociale reste le plus solide (résultats de l'AFC). Que pouvons-nous donc conclure de ces résultats ?

Comme nous pouvons le constater dans les résultats de l'AFC, la solution factorielle qui consiste en une double dimension, positive et négative (soit la solution retenue dans l'AFE), est également acceptable. De plus, lorsqu'on observe attentivement les résultats de l'AFE (rotation Varimax), on relève que trois énoncés sur cinq du facteur 2 présentent une saturation unique, les trois mêmes énoncés de la dimension sociale. Autrement dit, les deux autres énoncés qui présentent une double saturation pourraient être également sur la dimension individuelle. L'AFC a donc l'avantage de tester différentes solutions factorielles en permettant au chercheur de retenir la plus robuste ou encore de prendre connaissance du fait que plusieurs solutions factorielles « équivalentes » en théorie pourraient être alors considérées.

La théorie sous-jacente à l'estime de soi ainsi que les résultats empiriques d'études antérieures, mais aussi le contexte dans lequel l'ESST sera utilisé dans de futures études permettront au chercheur d'envisager la meilleure solution à utiliser en ayant toujours à l'esprit que la solution la plus robuste est celle qui présente une double dimension, individuelle et sociale.

Figure 22.3.
Résultats de l'analyse factorielle confirmatoire pour le modèle retenu



Légende : ESST : Estime de soi en tant que travailleur

Une autre interprétation possible de ces divergences entre les résultats de l'AFE et de l'AFC pourrait concerner le fait que l'échantillon testé avec l'AFE est anglophone alors que pour l'AFC, il est francophone. Ces différences linguistiques pourraient faire en sorte que les résultats soient

différents, quoique les caractéristiques des deux échantillons sont similaires sur de nombreux autres points (p. ex., vivre dans le même pays, avoir un diagnostic de trouble mental grave, être en processus de recherche d'emploi et, enfin, être inscrit dans un programme de soutien à l'emploi). À ce propos, lorsque des AFC avaient été effectuées auprès d'un autre échantillon anglophone de même nature, la solution factorielle la plus solide était celle qui présentait une double dimension, individuelle et sociale (Corbière *et al.*, 2009). Les différences linguistiques (et culturelles) peuvent cependant avoir un effet sur les résultats des réponses aux énoncés du questionnaire (voir le chapitre 24 de cet ouvrage).

CONCLUSION

Les analyses factorielles, exploratoire et confirmatoire, sont très populaires dans la communauté scientifique, car elles sont pour le chercheur un outil méthodologique intéressant, notamment pour réaliser la validation de questionnaires en sciences sociales, humaines et de la santé. L'analyse factorielle exploratoire (AFE) se situe en général en amont de celle qui est de nature confirmatoire, au sens où cette première cherche à faire émerger des facteurs dans lesquels des énoncés sont regroupés de façon homogène. Pour ce qui est de l'analyse factorielle confirmatoire (AFC), elle permet de confirmer ou d'infirmer ces modèles de mesure. Le point commun de ces deux analyses est sans conteste la recherche de la parcimonie. L'adoption de la loi de la parcimonie offre aux chercheurs comme aux cliniciens l'avantage d'utiliser des outils brefs.

Même si les AFE et AFC sont des outils méthodologiques fort intéressants et utiles pour la conception et la validation (p. ex., validité de construit) d'outils de mesure, il reste que le recrutement d'un bassin conséquent d'individus (>100) est parfois ardu, notamment dans le domaine de la santé. Il est donc important d'anticiper ce type d'écueils avant d'entreprendre une AFE ou une AFC. Aussi, toujours à titre de recommandation, il est préférable de travailler en équipe plutôt qu'en solo afin de discuter avec les autres membres de l'interprétation des résultats. Pour finir, il nous semble essentiel de souligner que même si plusieurs caractéristiques, coefficients et autres éléments statistiques relatifs à ces analyses ont été présentés dans ce chapitre, le lecteur est invité, d'une part, à consulter la bibliographie de ce chapitre pour approfondir ses connaissances et, d'autre part, à réaliser de nombreuses AFE et AFC pour se familiariser avec ce type d'analyse.

RÉFÉRENCES

Arbuckle, J.L. et W. Wothke (1999). *Amos 4.0 User's Guide*, Chicago, SPSS.

Asparouhov, T. et B. Muthén (2009). « Exploratory structural equation », *Structural Equation Modeling*, vol. 16, p. 397-438.

Babyak, M.A. et S.B. Green (2010). « Confirmatory factor analysis : An introduction for psychosomatic medicine researchers », *Psychosomatic Medicine*, vol. 72, nº 6, p. 587-597, <doi : 10.1097/PSY.0b013e3181de3f8a>.

Bacher, F. (1987). « Les modèles structuraux en psychologie. Présentation d'un modèle : Lisrel, première partie », *Le travail humain*, vol. 50, p. 347-370.

Bartlett, M.S. (1954). « A note on the multiplying factors for various chi square approximations », *Journal of the Royal Statistical Society*, Series A, Statistics in Society, 16(Series B), p. 296-298.

Beavers, A.S., J.W. Lounsbury, J.K. Richards, S.W. Huck, G.J. Skolits et S.L. Esquivel (2013). « Practical considerations for using exploratory factor analysis in educational research », *Practical Assessment, Research et Evaluation*, vol. 18, nº 6, p. 1-13.

Bentler, P.M. (1995). *EQS Structural Equations Program Manual*, Encino, Multivariate Software Inc.

Bentler, P.M. et D.G. Bonett (1980). « Significance tests and goodness of fit in the analysis of covariance structures », *Psychological Bulletin*, vol. 88, nº 3, p. 588-606, <doi : 10.1037/0033-2909.88.3.588>.

Bollen, K.A. (1989). *Structural Equations with Latent Variables*, New York, John Wiley and Sons.

Browne, M.W. (2001). « An overview of analytic rotation in exploratory factor analysis », *Multivariate Behavioral Research*, vol. 36, nº 1, p. 111-150, <doi : 10.1207/s15327906mbr3601_05>.

Bryant, F.B. et P.R. Yarnold (1995). « Principal components analysis and exploratory and confirmatory factor analysis », dans L.G. Grimm et R.R. Yarnold (dir.), *Reading and Understanding Multivariale Statistics*, Washington, D.C., American Psychological Association, p. 99-136.

Cattell, R.B. (1966). « The Scree Test for the number of factors », *Multivariate Behavioral Research*, vol. 1, nº 2, p. 245-276, <doi : 10.1207/s15327906mbr0102_10>.

Chou, C.P. et P.M. Bentler (1995). « Estimates and tests in structural equation modeling », dans R.H. Hoyle (dir.), *Structural Equation Modeling : Concepts, Issues, and Applications*, Thousand Oaks, Sage Publications, p. 37-54.

Comrey, A.L. et H.B. Lee (1992). *A First Course in Factor Analysis*, 2e éd., Hillsdale, Lawrence Erlbaum Associates.

Corbière, M., N. Lanctôt, N. Sanquirgo et T. Lecomte (2009). « Evaluation of self-esteem as a worker for people with severe mental disorders », *Journal of Vocational Rehabilitation*, vol. 30, p. 87-98.

Corbière, M., D. Reinharz, E. Latimer, B. Kirsh, C. Lecomte, P. Goering et E. Goldner (2006). *A Pan Canadian Analysis of Programmatic, Organizational, and Individual Aspects of Supported Employment Programs Implementation*, Ottawa, Canadian Institutes of Health Research (CIHR).

Costello, A. et J. Osborne (2005). « Best practices in exploratory factor analysis : Four recommendations for getting the most from your analysis », *Practical Assessment, Research et Evaluation*, vol. 10, p. 173-178, <doi : citeulike-article-id :7599621>.

Cudeck, R. et L.L. O'Dell (1994). « Applications of standard error estimates in unrestricted factor analysis : Significance tests for factor loadings and correlations », *Psychological Bulletin*, vol. 115, nº 3, p. 475-487.

Dassa, C. (1996). *Analyse multidimensionnelle exploratoire*, Montréal, Département de médecine sociale et prévention, Université de Montréal.

DiStefano, C. et B. Hess (2005). « Using confirmatory factor analysis for construct validation : An empirical review », *Journal of Psychoeducational Assessment*, vol. 23, nº 3, p. 225-241, <doi : 10.1177/073428290502300303>.

Fabrigar, L.R., D.T. Wegener, R.C. MacCallum et E.J. Strahan (1999). « Evaluating the use of exploratory factor analysis in psychological research », *Psychological Methods*, vol. 4, nº 3, p. 272-299, <doi : 10.1037/1082-989x.4.3.272>.

Gorsuch, R.L. (1983). *Factor Analysis*, 2e éd., Hillsdale, Erlbaum.

Guttmann, L.A. (1954). « A new approach to factor analysis : The radix », dans P.F. Lazarsfeld (dir.), *Mathematical Thinking in the Social Sciences*, New York, Columbia University Press.

Hair, J.F., W.C. Black, B.J. Babin, R.E. Anderson et R.L. Tatham (2006). *Multivariate Data Analysis*, 6e éd., New York, Macmillan.

Hendrickson, A.E. et P.O. White (1964). « Promax : A quick method for rotation to oblique simple structure », *British Journal of Statistical Psychology*, vol. 17, p. 65-70.

Henson, R.K. et J.K. Roberts (2006). « Use of exploratory factor analysis in published research : Common errors and some comment on improved practice », *Educational and Psychological Measurement*, vol. 66, nº 3, p. 393-416, <doi : 10.1177/0013164405282485>.

Hooper, D., J. Coughlan et M.R. Mullen (2008). « Structural equation modelling : Guidelines for determining model fit », *Electronic Journal of Business Methods*, vol. 6, nº 1, p. 53-60.

Hu, L.T. et P.M. Bentler (1999). « Cutoff criteria for fit indexes in covariance structure analysis : Conventional criteria versus new alternatives », *Structural Equation Modeling*, vol. 6, nº 1, p. 1-55, <doi : 10.1080/10705519909540118>.

Jackson, D.L., J.A. Gillaspy et R. Purc-Stephenson (2009). « Reporting practices in confirmatory factor analysis : An overview and some recommendations », *Psychological Methods*, vol. 14, nº 1, p. 6-23, <doi : 10.1037/a0014694>.

Jöreskog, K.G. et D. Sörbom (1993). *LISREL 8 : Structural Equation Modeling with the SIMPLIS Command Language*, Chicago, Scientific Software International Inc.

Jöreskog, K.G. et D. Sörbom (1996). *LISREL 8 : User's Reference Guide*, Lincolnwood, Scientific Software International Inc.

Kaiser, H. (1958). « The varimax criterion for analytic rotation in factor analysis », *Psychometrika*, vol. 23, nº 3, p. 187-200, <doi : 10.1007/bf02289233>.

Kaiser, H. (1970). « A second generation little jiffy », *Psychometrika*, vol. 35, nº 4, p. 401-415, <doi : 10.1007/bf02291817>.

Kaiser, H. (1974). « An index of factorial simplicity », *Psychometrika*, vol. 39, nº 1, p. 31-36, <doi : 10.1007/bf02291575>.

Lecomte, T., M. Corbière et F. Laisné (2006). « Investigating self-esteem in individuals with schizophrenia : Relevance of the Self-Esteem Rating Scale-Short Form », *Psychiatry Res*, vol. 143, nº 1, p. 99-108, <doi : 10.1016/j.psychres.2005.08.019>.

Ledesma, R.D. et P. Valero-Mora (2007). « Determining the number of factors to retain in EFA : An easy-to-use computer program for carrying out parallel analysis », *Practical Assessment, Research and Evaluation*, vol. 12, nº 2, p. 1-11

Lent, R.W., S.D. Brown et P.A.J. Gore (1997). « Discriminant and predictive validity of academic self-concept, academic self-efficacy, and mathematics-specific self-efficacy », *Journal of Counseling Psychology*, vol. 44, nº 3, p. 307-315.

Mueller, R.O. (1996). *Basic Principles of Structural Equation Modeling : An Introduction to LISREL and EQS*, New York, Springer-Verlag.

Muthén, L.K. et B. Muthén (1998). *Mplus User's Guide. Version 6*, Los Angeles, Muthén and Muthén.

Norušis, M.J. et SPSS Inc. (2003). *SPSS 12.0 : Statistical Procedures Companion*, Upper Saddle, Prentice Hall PTR.

Norušis, M.J. et SPSS Inc. (2005). *SPSS 13.0 : Statistical Procedures Companion*, Upper Saddle, Prentice Hall PTR.

O'Brien, K. (2007). « Factor analysis : An overview in the field of measurement », *Physiotherapy Canada*, vol. 59, nº 2, p. 142-155, <doi : 10.3138/ptc.59.2.142>.

O'Connor, B.P. (2000). « SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test », *Behavior Research Methods, Instruments, and Computers*, vol. 32, p. 396-402

Pedhazur, E.J. et L.P. Pedhazur Schmelkin (1991). *Measurement, Design, and Analysis : An Integrated Approach*, Hillsdale, Erlbaum.

Pett, M.A., N.R. Lackey et J.J. Sullivan (2003). *Making Sense of Factor Analysis – The Use of Factor Analysis for Instrument Development in Health Care Research*, Thousand Oaks, Sage Publications.

Quintana, S.M. et S.E. Maxwell (1999). « Implications of recent developments in structural equation modeling for counseling psychology », *The Counseling Psychologist*, vol. 27, nº 4, p. 485-527, <doi : 10.1177/0011000099274002>.

Ritter, N.L. (2012). « A comparison of distribution free and non-distribution free factor analysis methods », communication présentée au congrès annuel de la Southwest Educational Research Association, Nouvelle-Orléans.

Rosenberg, M. (1965). *Society and the Adolescent Self-Image*, Princeton, Princeton University Press.

Russell, D.W. (2002). « In search of underlying dimensions : The use (and abuse) of factor analysis in personality and social psychology bulletin », *Personality and Social Psychology Bulletin*, vol. 28, nº 12, p. 1629-1646, <doi : 10.1177/014616702237645>.

Schermelleh-Engel, K., H. Moosbrugger et H. Müller (2003). « Evaluating the fit of structural equation models : Tests of significance and descriptive goodness-of-fit measures », *Methods of Psychological Research*, vol. 8, nº 2, p. 23-74.

Schmitt, T.A. (2011). « Current methodological considerations in exploratory and confirmatory factor analysis », *Journal of Psychoeducational Assessment*, vol. 29, nº 4, p. 304-321, <doi : 10.1177/0734282911406653>.

Spearman, C. (1904). « "General Intelligence", objectively determined and measured », *The American Journal of Psychology*, vol. 15, nº 2, p. 201-292, <doi : 10.2307/1412107>.

Tabachnick, B.G. et L.S. Fidell (1996). *Using Multivariate Statistics*, 3e éd., New York, HarperCollins College Publishers.

Tabachnick, B.G. et L.S. Fidell (2001). *Using Multivariate Statistics*, 4e éd., Needham Heights, Allyn et Bacon.

Tabachnick, B.G. et L.S. Fidell (2007). *Using Multivariate Statistics*, 6e éd., Londres, Pearson Education.

Thompson, B. (2004). *Exploratory and Confirmatory Factor Analysis : Understanding Concepts and Applications*, Washington, American Psychological Association.

Thurstone, L.L. (1938). *Primary Mental Abilities (Psychometric Monograph No. 1)*, Chicago, University of Chicago Press.

CHAPITRE 23

LA CARTOGRAPHIE DE CONCEPTS
Une représentation visuelle et spatiale pour décrire les ressources résidentielles en santé mentale

Amélie Felx
Mary Kane
Marc Corbière
Alain Lesage

FORCES

- Elle facilite la compréhension d'un concept abstrait ou d'une intervention complexe par l'intermédiaire d'une représentation visuelle et spatiale.
- Elle permet de recueillir et d'analyser les perceptions divergentes de multiples parties prenantes.
- Elle consiste en une technique structurée qui peut être utilisée auprès de petits comme de grands groupes de participants.

LIMITES

- Le processus peut être exigeant pour les participants (et le chercheur).
- L'interprétation des cartes conceptuelles produites peut, dans certains cas, s'avérer laborieuse.
- La validité externe peut être limitée en raison de la représentativité (ou non) des participants.

La cartographie de concepts[1] ou le *concept mapping* désigne le processus par lequel des idées sont représentées sous la forme de cartes conceptuelles. La cartographie de concepts est plus vaste que la technique décrite dans ce chapitre et développée par William M.K. Trochim de l'Université Cornell (Kane et Trochim, 2007; Trochim, 1989). En effet, plusieurs variantes en existent, et leur utilisation s'étend au-delà de l'évaluation et de la recherche. Par exemple, elle peut servir pour la planification stratégique, pour favoriser les apprentissages et parfois même pour les évaluer (Markham, Mintzer et Jones, 1994; Novak et Gowin, 1984).

En tant que méthode ou technique de collecte et d'analyse de données, la cartographie de concepts représente une option parmi d'autres pour explorer des questions entourant les interventions complexes, la définition des besoins, la théorie des programmes, les priorités d'une organisation, etc. Son utilisation dans le domaine de la santé mentale est peu fréquente comparativement à celle d'autres techniques qui peuvent aussi servir à structurer la réflexion et la communication au sein de groupes de participants, telles que la méthode Delphi (voir le chapitre 13 de cet ouvrage), la méthode TRIAGE (voir le chapitre 14 de cet ouvrage), le groupe de discussion focalisée (*focus group*) (voir le chapitre 12 de cet ouvrage), la technique du groupe nominal et le *brainwriting* (Pinault et Daveluy, 1986).

Ce chapitre comprend deux parties. La première décrit les six étapes de la cartographie de concepts (Kane et Trochim, 2007)[2]. La deuxième présente le déroulement et les résultats d'une étude multicentrique québécoise dans laquelle la cartographie de concepts a été utilisée pour décrire les ressources résidentielles en santé mentale. À partir de cet exemple d'application, quelques-uns des défis rencontrés par l'équipe de recherche sont mis en relief.

1. CARTOGRAPHIE DE CONCEPTS

1.1. Caractéristiques générales

La cartographie de concepts consiste en un processus de conceptualisation structurée (*structured conceptualization*) comprenant six étapes illustrées à la figure 23.1 de la section 1.3. Elle permet de déterminer les principales

1. Dans ce chapitre, l'expression «cartographie de concepts» est une traduction de l'anglais *concept mapping*. «Cartographie conceptuelle» est aussi utilisée par certains auteurs.
2. Ce chapitre rédigé en collaboration avec l'un des concepteurs de la technique (M. Kane) en reprend les principaux éléments décrits dans les textes clés (Kane et Trochim, 2007; Trochim, 1989).

composantes, dimensions ou particularités d'une réalité ou d'un phénomène et d'en visualiser l'univers conceptuel sous-jacent. Le tableau 23.1 présente quelques-unes de ses particularités.

Tableau 23.1.
Particularités de la cartographie de concepts

- Repose sur une collecte systématique d'informations et sur des étapes bien définies.
- Permet de regrouper des parties prenantes hétérogènes avec des points de vue et des intérêts divergents.
- Favorise la participation des parties prenantes aux différentes étapes de la démarche (approche participative).
- Recueille et analyse des données qualitatives et quantitatives, incluant des analyses multivariées (approche mixte).
- Utilise habituellement plusieurs outils de collecte de données tels que le questionnaire, le remue-méninges, la réflexion individuelle, le groupe de discussion focalisée, etc.
- Ne vise pas nécessairement la recherche d'un consensus.

La cartographie de concepts permet de regrouper une hétérogénéité de parties prenantes avec des préoccupations propres à différents niveaux hiérarchiques et qui ont une expérience vécue ou théorique en lien avec le sujet ou le phénomène étudié (p. ex., gestionnaires, intervenants, personnes utilisatrices de services). Elle confère un poids équivalent aux perceptions de chaque participant et utilise leur langage. La participation des parties prenantes prend diverses formes aux différentes étapes de la démarche, sauf lors de l'analyse des données, qui est réalisée par le chercheur (étape 4 ; voir section 1.3.4). La nature des interactions entre les participants varie aussi selon les étapes de la démarche et selon les choix méthodologiques faits par le chercheur. Habituellement, la cartographie de concepts tend à limiter le degré de liberté et d'ouverture dans les échanges entre les participants, ce qui permet de restreindre l'effet d'entraînement du discours d'une personne sur une autre (voir section 1.3.2).

La cartographie de concepts est qualifiée d'approche mixte, car ses composantes qualitatives et quantitatives sont intrinsèquement rattachées (Kane et Trochim, 2007). Les données qualitatives recueillies sont analysées par l'intermédiaire d'analyses qualitatives (analyse de contenu ou textuelle) et quantitatives. Ces dernières comprennent des statistiques descriptives et des analyses multivariées, soit l'échelonnage multidimensionnel (*multidimensional scaling*) et la typologie hiérarchique (*hierarchical cluster analysis*) (voir le chapitre 15 de cet ouvrage). Ce fondement mathématique et le

recours à l'algèbre matricielle sont des particularités de la technique. La plupart des autres variantes des cartes conceptuelles ne reposent pas sur des analyses statistiques (p. ex., Novak et Gowin, 1984).

Les résultats de la cartographie de concepts prennent la forme d'une représentation visuelle et spatiale (cartes conceptuelles). Ces cartes synthétisent une grande quantité d'informations et sont plus aisément compréhensibles que les tableaux de résultats plus traditionnels en recherche (p. ex., les tableaux des indices de saturation des items à un facteur issus de l'analyse factorielle) (voir le chapitre 22 de cet ouvrage). Les cartes conceptuelles tiennent compte de l'ensemble des différentes perceptions individuelles et mettent en relation les composantes ou dimensions du phénomène étudié. Dans les faits, la cartographie de concepts n'a pas comme premier objectif d'obtenir un consensus entre les participants; elle vise plutôt à recueillir une grande diversité de perceptions pour explorer en profondeur le sujet ou le phénomène d'intérêt. La recherche de consensus entre les participants est remplacée par l'agrégation des données. La cartographie de concepts permet donc d'aborder les questions de recherche sous l'angle de la pluralité (c.-à-d. parties prenantes, réalités, savoirs). Selon les valeurs et les choix du chercheur (voir section 1.3), la démarche pourra s'inscrire dans un des grands paradigmes existants[3].

1.2. Évolution et utilisation de la cartographie de concepts

Depuis les premières publications dans les années 1980 à son sujet, la cartographie de concepts s'est raffinée et son utilisation s'est répandue (Rosas et Kane, 2012). Plus de 120 travaux de maîtrise et de doctorat s'y rapportent, et plus de 200 articles scientifiques publiés décrivent des projets qui utilisent la cartographie de concepts ou des adaptations de la technique (Johnsen, Biegel et Shafran, 2000). Le lecteur intéressé peut se renseigner sur le développement de la cartographie de concepts et retracer plusieurs publications et exemples de projets sur le site Web <http://www.conceptsystems.com>. Sauf quelques exceptions (p. ex., Trochim, 1993; Trochim, Cook et Setze, 1994; Jackson et Trochim, 2002; Southern *et al.*, 2002; Rosas, 2005; Dagenais *et al.*, 2009; Rosas et Kane, 2012), peu de publications font état des limites, de la qualité des représentations obtenues et de l'applicabilité de la technique. Ces exceptions permettent d'amorcer une réflexion autour de la fiabilité et de la validité de la cartographie de concepts et fournissent des pistes pour juger de la qualité d'un projet.

3. En ce sens, Kane et Trochim (2007) parlent d'une ontologie neutre (p. 176).

1.2.1. Domaines de recherche et exemples d'utilisation

Les finalités ou les utilisations possibles de la cartographie de concepts sont multiples : par exemple, la planification, l'élaboration d'un cadre conceptuel ou d'un instrument de mesure, la définition de besoins, la conception d'une intervention ou d'un programme, ou le transfert de connaissances. En revanche, elle est très rarement utilisée pour inférer la causalité ou tester une hypothèse.

La cartographie de concepts a été abondamment utilisée dans le domaine de la planification et de l'évaluation. Plusieurs revues d'évaluation ont publié des articles portant sur des projets l'ayant mise à contribution (p. ex., *Evaluation and Program Planning*, *American Journal of Evaluation*). La santé publique et la santé communautaire sont les disciplines dans lesquelles la cartographie de concepts est la plus fréquemment utilisée (Rosas et Kane, 2012). Elle est aussi utilisée en sciences sociales, en santé et en sciences biomédicales (Rosas et Kane, 2012). Quelques publications se rapportent à la santé mentale et à la réadaptation (Johnsen *et al.*, 2000). Par exemple, elle a servi à créer un cadre conceptuel et à établir les composantes d'un programme de soutien à l'emploi (Trochim *et al.*, 1994), et à explorer les facteurs qui influencent l'adhésion au traitement médicamenteux telle que perçue par les intervenants, les proches et les personnes atteintes de schizophrénie (Kikkert *et al.*, 2006).

1.2.2. Logiciels

Les logiciels d'analyses statistiques courants (p. ex., SPSS et SAS) peuvent être utilisés pour réaliser les analyses quantitatives menant à la création de cartes conceptuelles. Leur utilisation requiert une connaissance approfondie des analyses multivariées. En comparaison, le prologiciel Concept System® (<http://www.conceptsystems.com/content/view/the-concept-system.html>), développé par William M.K. Trochim spécialement pour la technique, exécute automatiquement chacune des transformations et des opérations mathématiques nécessaires à la production des résultats et des cartes conceptuelles. Le prologiciel comprend aussi une interface Web qui permet aux participants d'y entrer directement leurs réponses aux étapes 2 et 3 de la démarche (sections 1.3.2. et 1.3.3). De plus, une équipe d'experts est disponible pour accompagner le chercheur dans sa démarche et dans l'utilisation du prologiciel. Des formations en ligne et des cours sont aussi offerts. Enfin, des chercheurs de Concept System® travaillent depuis plusieurs années au raffinement du prologiciel. La nouvelle version du prologiciel (2012) est plus flexible et permet l'exportation de certaines données vers d'autres logiciels (p. ex., SPSS, Excel). Son coût a aussi été réduit.

1.3. Cartographie de concepts : étape par étape

Conçue comme une démarche systématique et rigoureuse en six étapes, la cartographie de concepts laisse tout de même de la place à une certaine flexibilité (figure 23.1). Les particularités de chaque projet seront intrinsèquement reliées au sujet ou au phénomène d'intérêt, au contexte, au paradigme d'appartenance du chercheur, à la distribution géographique des participants, à la diversité des parties prenantes ainsi qu'au temps et aux ressources disponibles. Entre autres, le chercheur peut choisir parmi différentes techniques de collecte de données ou opter pour celles traditionnellement utilisées dans les projets de cartographie de concepts pour générer les idées (remue-méninges), les coter (questionnaire) et les organiser (classification libre) (Rosenberg et Kim, 1975). Kane et Trochim (2007) proposent des techniques alternatives qui ne sont pas abordées dans ce chapitre[4].

1.3.1. Préparer la démarche

Le travail préparatoire représente une étape cruciale de la cartographie de concepts ; les décisions prises à ce moment influencent le déroulement de la démarche et sa qualité. Kane et Trochim (2007) décrivent huit sous-étapes qui permettent de bien préparer la démarche. Les principales sous-étapes sont décrites brièvement ci-dessous. Toutefois, avant toute chose, il importe de réfléchir sur le bien-fondé d'utiliser la cartographie de concepts et de s'assurer que la problématique et les objectifs de la démarche sont clairs.

Le choix du modérateur

Le choix d'un modérateur est un enjeu dans tout processus de groupe ; il est aussi un facteur déterminant du succès ou non de la procédure. Selon le contexte et l'objectif de la démarche, différentes qualités peuvent être recherchées chez le modérateur (p. ex., interne ou externe au programme, convivialité, expérience, capacité d'encadrement).

*La formulation de l'énoncé pour le remue-méninges (*focus*)*

L'énoncé qui sera présenté aux participants à l'étape 2 peut prendre la forme d'une consigne, d'une question ou d'une phrase à compléter (tableau 23.2). La phrase à compléter a l'avantage de générer des idées avec une structure grammaticale plus uniforme.

4. Par exemple, pour générer des idées (étape 2), il est possible de recourir à un groupe de discussion focalisée, à des entrevues ou à la littérature. Pour la cotation (étape 3), il est possible de demander aux participants de ressortir les dix idées les plus importantes.

Figure 23.1.
Les six étapes de la cartographie de concepts

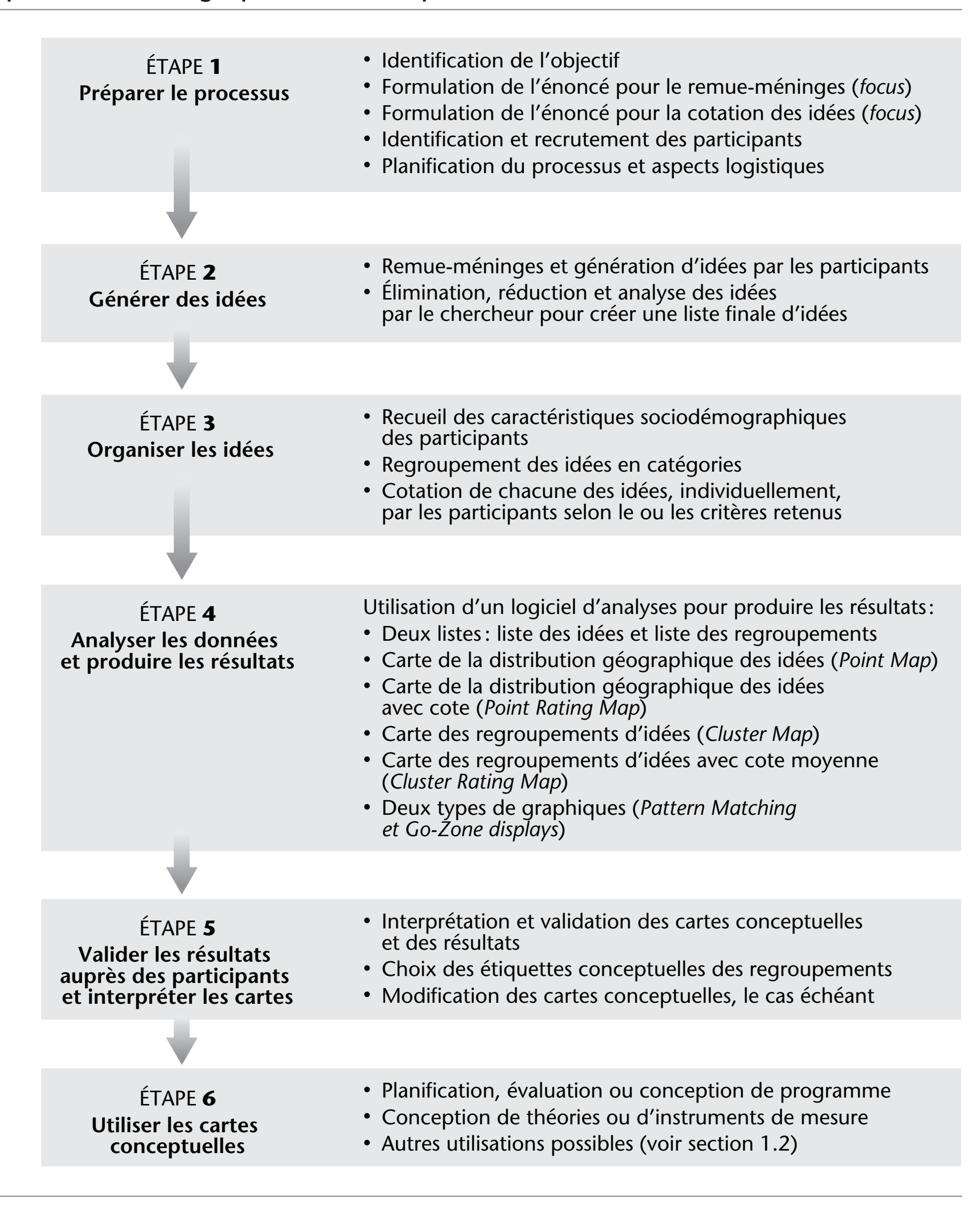


Tableau 23.2.
Exemples d'énoncés pour un remue-méninges

1. Consigne	Générez de courtes phrases qui décrivent les besoins de formation des professionnels œuvrant dans l'unité de traitement et de réadaptation intensive.
2. Question	Quels sont les besoins de formation des professionnels œuvrant dans l'unité de traitement et de réadaptation intensive ?
3. Phrase à compléter	Les professionnels de l'unité de traitement et de réadaptation intensive auraient un besoin de formation sur…

L'énoncé est à la base du remue-méninges qui recueille l'ensemble des idées qui serviront à la construction des cartes conceptuelles. Sa formulation doit être claire et univoque. L'énoncé doit permettre de générer des idées qui couvrent l'ensemble du phénomène ou du domaine conceptuel ciblé tout en favorisant l'émergence d'idées créatives. En contrepartie, l'énoncé doit être suffisamment précis pour limiter les idées hors sujet. Le lecteur intéressé à la formulation d'un énoncé peut consulter les écrits sur cette question (p. ex., Fowler, 2009).

*La formulation de l'énoncé pour la cotation des idées (*focus*)*

Afin de préparer l'étape 3, où les participants devront attribuer une valeur numérique aux idées émises lors du remue-méninges, le chercheur doit au cours de cette sous-étape déterminer sur la base de quel(s) critère(s) chaque participant cotera chacune des idées (voir section 1.3.3). Cette cotation permettra de mettre les différentes idées en lien les unes avec les autres. N'importe quelle mesure d'intérêt peut servir de critère : importance, coût, priorité, faisabilité, etc. (voir partie 2 pour un exemple).

La sélection et le recrutement des participants

L'identification des participants est intrinsèquement liée au contexte et à l'objectif de la démarche. Il faut déterminer le nombre et le type de participants qui prendront part aux différentes étapes de la démarche (p. ex., intervenants, personnes utilisatrices de services, gestionnaires, personnel administratif). Leur sélection peut reposer sur un échantillonnage aléatoire si le chercheur veut un groupe de participants représentatif d'une population cible, ou sur un échantillonnage par choix raisonné s'il souhaite obtenir une diversité de perspectives. Il n'est pas nécessaire que les mêmes participants soient impliqués dans chacune des étapes de la démarche (Kane et Trochim, 2007 ; Johnsen *et al.*, 2000). Toutefois, le fait de ne pas participer à l'ensemble de la démarche peut nuire à la compréhension qu'en

ont les participants. Enfin, une fois les participants potentiels sélectionnés, il faut procéder à leur recrutement et, le cas échéant, à l'obtention de leur consentement libre et éclairé.

La recension de 69 projets utilisant la cartographie de concepts réalisés entre 2001 et 2011 (Rosas et Kane, 2012) indique que le nombre moyen de participants est de 155,78 (étendue : 20 à 649). Les projets qui se déroulent sur le Web comptent en moyenne plus de participants (243,42). Bien qu'il n'existe aucun critère précis pour déterminer le nombre total de participants souhaité, pour la classification des idées (étape 3), un nombre entre 20 et 30 maximisera la valeur de l'indice de stress (Rosas et Kane, 2012). Cet indice permet de juger de la qualité de l'ajustement du modèle final et de déterminer si la représentation obtenue est aléatoire ou non.

La méthode

Il s'agit ici de préciser plusieurs des éléments qui viendront établir les paramètres de la démarche. Par exemple, le chercheur décide du ou des modes de collecte de données (p. ex., rencontres de groupe en face à face, courrier postal ou électronique, Web, modes multiples). Il doit aussi répondre à plusieurs questions comme : « Dans quelle fenêtre temporelle se déroulera la démarche : plusieurs mois, trois semaines, une journée ? » « Y aura-t-il un comité consultatif qui accompagnera le chercheur dans les différentes étapes de la démarche[5] ? »

En somme, à l'étape 1, la démarche se dessine et le plan de travail détaillé est rédigé.

1.3.2. Générer des idées

En général, les participants sont invités à prendre part à une séance de remue-méninges, au cours de laquelle ils sont appelés à formuler de courtes idées ou phrases en réponse à l'énoncé formulé à l'étape 1. Comme mentionné précédemment, le remue-méninges peut être réalisé lors de une ou plusieurs rencontres de groupe, par courrier postal ou électronique, ou sur le Web. Les rencontres de groupe requièrent planification et organisation (tableau 23.3) ; toutefois, elles ont l'avantage d'augmenter le taux de participation, de générer plus d'interactions entre les participants et de les inciter à se centrer sur la tâche. Le Web, en contrepartie, permet de rejoindre facilement et rapidement des participants dispersés sur un territoire, et est moins contraignant pour les participants.

5. La présence d'un comité consultatif est recommandée par Kane et Trochim (2007) et rend la démarche plus objective et plus participative.

Tableau 23.3.
Étapes de la préparation d'une rencontre de remue-méninges

Participants	Envoyer des invitations et de l'information sur le projet et son contexte. Envoyer ou faire un rappel aux participants avant la rencontre.
Modérateur	Préparer le modérateur (p. ex., préciser les attentes, expliquer les enjeux). Préparer le texte et les lignes directrices pour le modérateur.
Lieu de la rencontre	Réserver une salle de rencontre de qualité (p. ex., confort, luminosité). S'assurer de la bonne disposition de la salle (p. ex., en rectangle). S'assurer que tous les participants verront bien l'écran.
Matériel	S'assurer d'avoir le matériel nécessaire (p. ex., ordinateur, projecteur, écran, papier et crayons, tableau, documents sur le projet, formulaires de consentement). Prévoir un plan B si l'ordinateur ne fonctionne pas. Offrir café et viennoiseries.
Contenu	Présenter le contexte et la démarche. Préciser le temps consacré aux différentes tâches (p. ex., pause).
Déroulement	Réaliser un prétest et ajuster la démarche en conséquence.

Source : Adapté de Kane et Trochim, 2007.

Une rencontre de groupe débute, en général, par une brève présentation du projet et de ses objectifs. Les règles et les principes du remue-méninges doivent aussi être expliqués (Osborn, 1948). Le remue-méninges se fonde sur deux principes : la suspension du jugement et la recherche la plus étendue d'idées possibles tout en respectant l'énoncé. Les participants peuvent poser des questions afin de clarifier le sens d'une idée. Toutefois, il faut éviter les discussions et les commentaires autour de la valeur des idées émises. La critique peut conduire à l'échec de la procédure.

Une fois l'énoncé posé, les participants prennent quelques minutes pour y réfléchir. Par la suite, chaque participant formule une idée (ou choisit de passer son tour). Les tours de table se poursuivent jusqu'à l'épuisement des idées. Le modérateur peut, par la suite, suggérer une période où des idées sont émises en vrac. Le modérateur voit à ce que chacun des participants ait la parole et à ce qu'il n'y ait aucune critique (explicite ou implicite) des idées exprimées. Il doit aussi s'assurer de la fluidité des idées sans en influencer la production. Les idées émises sont inscrites au fur et à mesure sur un tableau ou transcrites sur un ordinateur rattaché à un écran projecteur. Les participants peuvent donc visualiser simultanément l'ensemble des idées émises par le groupe et en valider la saisie.

À la suite de la génération des idées, un travail d'édition est habituellement nécessaire pour créer une liste finale d'idées qui sera présentée aux participants à l'étape 3. Les idées redondantes sont regroupées, certaines idées sont reformulées et les idées non pertinentes sont éliminées. Parfois, le nombre d'idées générées est tel qu'un travail d'analyse de contenu ou textuelle est souhaitable. Ce travail d'édition et d'analyse est primordial, car les idées doivent être claires et représenter la variété et la richesse des idées énoncées par les participants. L'homogénéité de leur structure grammaticale facilitera le travail des participants à l'étape suivante. Au demeurant, plus le nombre d'idées est élevé, plus la tâche des participants sera complexe (étape 3). Kane et Trochim (2007) recommandent que ce nombre soit inférieur à 100 et, selon la recension de Rosas et Kane (2012), les participants génèrent en moyenne 93,32 idées (é.t. = 17,23).

1.3.3. Organiser (classifier et coter) les idées

Une fois la liste d'idées créée, les participants réalisent, individuellement, deux tâches : 1) regrouper les idées ; 2) coter les idées selon un ou plusieurs critères. À cette étape, il importe que le travail produit individuellement puisse être rattaché à un participant. Pour ce faire, des données descriptives sont recueillies (p. ex., numéro du participant, genre, âge). Ces données permettront d'effectuer des comparaisons entre les réponses des différents participants.

La première tâche consiste à regrouper les idées en catégories ou en regroupements (Rosenberg et Kim, 1975 ; Weller et Romney, 1988). Chaque participant peut former autant de regroupements qu'il le désire, selon une logique de classification qu'il déterminera et en faisant appel à des schèmes qui représentent sa façon de conceptualiser le sujet ou le phénomène d'intérêt. Le nombre de regroupements peut aller de 1 (toutes les idées sont regroupées ensemble en un seul regroupement) au nombre total d'idées (chaque idée constitue un regroupement). Trois consignes sont données aux participants : 1) ne pas mettre une même idée dans plus d'un regroupement ; 2) ne pas laisser toutes les idées seules (sans les regrouper) ; 3) ne pas regrouper les idées qui semblent ne se rattacher à aucun regroupement (il est préférable de laisser l'idée seule).

Pour ce faire, le participant révise l'ensemble des idées. Ensuite, pour chaque idée, le participant se questionne. Par exemple, à la figure 23.2, si l'un des quatre participants tente de classer l'idée 7 (i7), il se demandera si cette idée rejoint les idées 2 (i2), 5 (i5) et 3 (i3) quant à sa nature, sa signification ou tout autre attribut. Si oui, le participant place l'idée 7 (i7) dans le regroupement R1. Sinon, le participant poursuit son questionnement :

l'idée 7 (i7) rejoint-elle les idées du regroupement R2 ? Si l'idée 7 (i7) ne rejoint aucun des deux regroupements existants, elle formera un troisième regroupement (R3). Le participant poursuit avec l'idée suivante, jusqu'à ce que toutes les idées soient placées. Pour chaque regroupement formé, le participant indique une appellation ou la raison pour laquelle il a regroupé les idées ensemble. Si la procédure se déroule en groupe, chaque idée est imprimée sur un carton individuel et chaque participant crée des regroupements (piles de cartons) manuellement.

La deuxième tâche consiste à attribuer une cote à chacune des idées, selon le ou les critère(s) retenu(s) à l'étape 1. Par exemple, si le critère retenu est l'importance, chaque participant cote l'importance de chacune des idées. Différentes échelles de cotation peuvent être utilisées, la plus fréquente étant celle de Likert (généralement entre quatre et sept points). Si la procédure se déroule en groupe, le modérateur remet à chaque participant un questionnaire sur lequel il pourra encercler ses réponses.

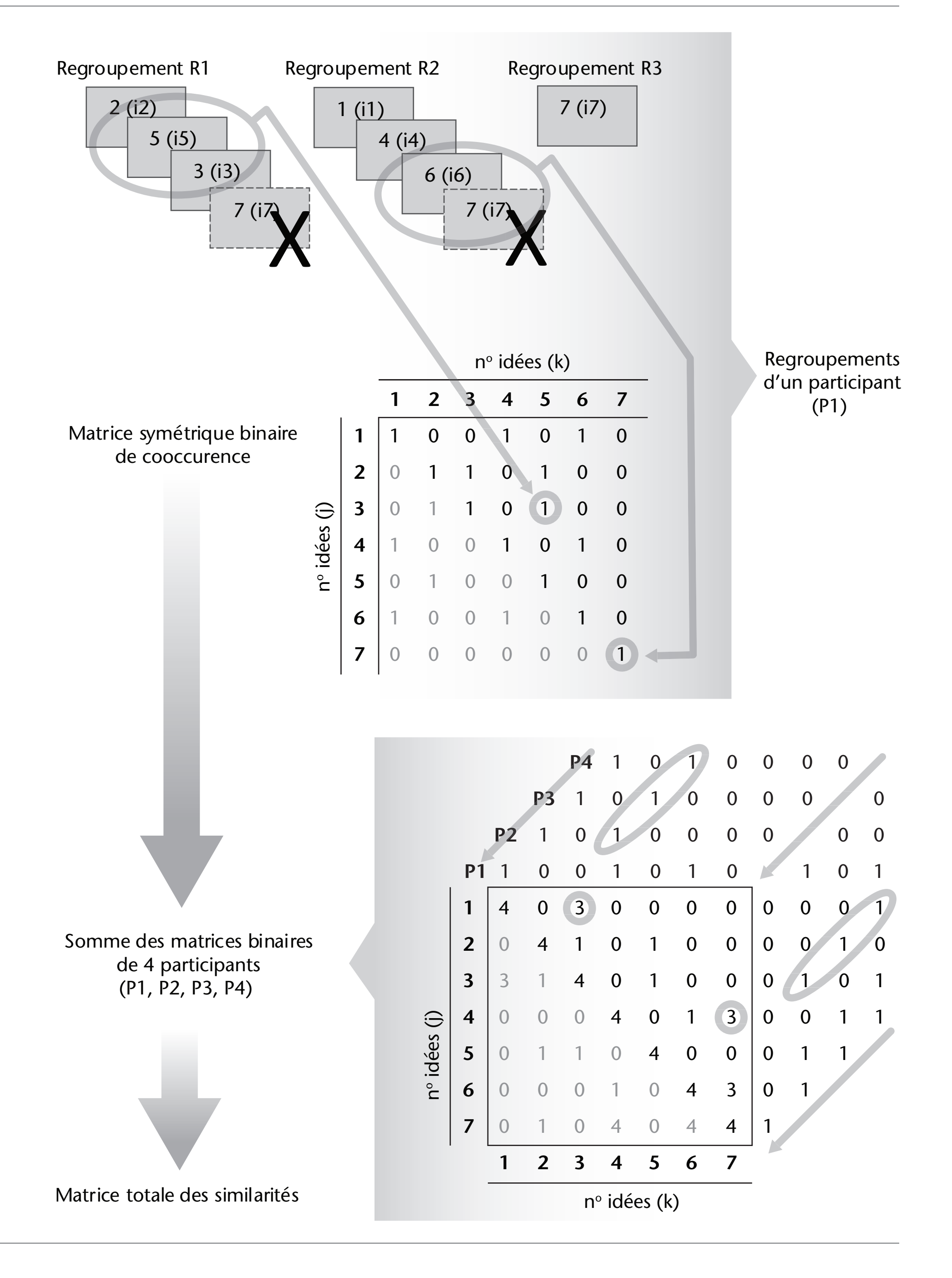
1.3.4. Analyser les données et créer les cartes conceptuelles

Les résultats de la cartographie de concepts découlent d'une série d'analyses de données[6] : 1) l'échelonnage multidimensionnel ; 2) la typologie hiérarchique ; 3) le calcul de moyennes ; et 4) le calcul de corrélations. Les résultats produisent des listes, deux types de graphiques et quatre types de cartes. Ils sont décrits ci-dessous et plusieurs sont illustrés dans la partie 2 du chapitre (figures 23.3 à 23.6).

L'échelonnage multidimensionnel recherche une structure dans un espace affine (issu de la géométrie euclidienne) et repose sur la transformation des données qualitatives recueillies en matrices (figure 23.2). Dans l'exemple fictif illustré à la figure 23.2, un participant a créé trois regroupements à partir de sept idées. Si deux idées (j et k) sont regroupées par un participant, le chiffre 1 apparaît dans la matrice résultante ($r_{j,k}$). Si deux idées (j et k) ne sont pas regroupées, le chiffre 0 apparaît ($r_{j,k}$). Les matrices symétriques binaires des productions individuelles sont additionnées pour former une matrice symétrique totale de similarités (*total square similarity matrix*), qui permet de calculer un indice de proximité ou de distance entre j et k. Les nombres dans cette deuxième matrice indiquent le nombre de participants qui ont regroupé les deux mêmes idées ensemble. À la figure 23.2, un 4 indique que tous les participants ($n = 4$) ont regroupé deux idées (j et k)

6. Les détails des analyses statistiques dépassent le cadre de ce chapitre. Le lecteur intéressé peut se référer à des ouvrages sur les méthodes multivariées ou à des ouvrages portant précisément sur l'échelonnage multidimensionnel et la typologie hiérarchique (p. ex., Kruskal et Wish, 1978 ; Anderberg, 1973 ; Everitt, 1980).

Figure 23.2.
Exemple de regroupements d'idées et leurs transformations en matrices



ensemble. Ainsi, plus ce nombre est élevé, plus un nombre élevé de participants font l'association entre les deux idées, ce qui laisse envisager une proximité conceptuelle (ou une probabilité élevée d'être effectivement reliées conceptuellement). À partir de cette information, l'échelonnage multidimensionnel permet la création d'une carte de la distribution géographique des idées (*point map*), qui représente visuellement la proximité ou la distance entre les idées émises dans un espace à dimensions réduites (figure 23.3). Le nombre de dimensions est fixé à deux par le prologiciel, et ce, comme justifié par Rosas et Kane (2012) et par Kruskal et Wish (1978), lorsque l'objectif est de créer des regroupements (*clusters*). De plus, l'interprétation des résultats est plus simple dans un espace à deux dimensions. L'algorithme fixe une géométrie non métrique, c'est-à-dire un mode pour calculer des distances entre les points et pour représenter les relations de proximité observées entre les idées dans l'espace euclidien. Sur la carte résultante, chaque point correspond à une idée et occupe des coordonnées (X_j Y_j). Les idées les plus fortement associées se situent plus près les unes des autres.

La deuxième série d'opérations mathématiques repose sur l'analyse typologique hiérarchique (*hierarchical cluster analysis*) et utilise les coordonnées (X_j Y_j) issues de l'échelonnage multidimensionnel. Elle permet de regrouper les idées en classes ou grappes exclusives (*clusters*) (voir le chapitre 15 de cet ouvrage), de telle sorte que les idées d'un même regroupement soient les plus semblables possibles et représentent un concept similaire. L'algorithme de Ward (Everitt, 1980) utilise une classification ascendante et permet de produire n'importe quel nombre de grappes (allant de 1 au nombre total d'idées) de façon itérative. Lors de chaque étape, l'algorithme agrège deux regroupements formés à l'étape précédente, créant ainsi une arborisation. Il n'existe aucun critère statistique pour déterminer le nombre idéal de regroupements. Le chercheur doit procéder à un examen des différentes solutions et il a la tâche, à la fois difficile et passionnante, de rechercher une solution optimale à partir de critères tels que la parcimonie, le potentiel d'interprétation, le contenu des regroupements et la perte d'informations. Une fois le nombre de regroupements déterminé, la carte des regroupements d'idées (*cluster map*) est produite (figure 23.4). À ce stade, la solution n'est pas finale, car les participants pourront la modifier à l'étape 5.

La troisième série d'opérations mathématiques consiste à calculer la moyenne des cotes attribuées à chacune des idées par les participants pour les critères retenus. Une fois ces opérations réalisées, deux cartes sont produites : la carte de la distribution des points avec cote (*point rating map*) et la carte des regroupements avec cote (*cluster rating map*) (figure 23.5). Le prologiciel permet aussi la création de deux types de graphiques additionnels

qui illustrent les relations entre les cotes moyennes attribuées aux idées ou aux regroupements et les caractéristiques des participants. Le premier consiste en un nuage de points (idées) dans un système d'axes (X, Y) (*go-zone display*) (figure 23.6). Le deuxième met en relation les cotes moyennes des regroupements en fonction de deux attributs (critères ou caractéristiques des participants) (*pattern matching display)* (figure 23.6). Un coefficient de corrélation de Pearson est aussi calculé.

Tous les résultats produits reflètent différentes façons de représenter le phénomène d'intérêt et de mieux le comprendre. À cela s'ajoutent différents indices (p. ex., itérations, indice de dispersion) et moyens (p. ex., indice de stress de Kruskal, méthode des deux moitiés) sur lesquels peut s'appuyer l'interprétation des résultats et qui peuvent être utilisés pour explorer la fiabilité de la démarche et sa validité (Jaworska et Chupetlovska-Anastasova, 2009 ; Rosas et Kane, 2012).

1.3.5. Valider les résultats et interpréter les cartes conceptuelles

À cette étape, il y a un retour auprès des participants pour faire ressortir, de façon collective, le sens qui émerge des résultats et des cartes conceptuelles. Pour la première fois dans la démarche, des discussions ont lieu entre les participants, et le groupe tente habituellement d'en arriver à un consensus sur le sens qui émerge de cette représentation. Lors de la rencontre, les différents résultats sont présentés aux participants, qui examinent les idées comprises dans chacun des regroupements. Bien que cela soit déconseillé par les experts de la technique, le groupe peut décider de déplacer une (ou quelques idées) vers un regroupement auquel elle semble mieux se rattacher et, par le fait même, modifier les résultats obtenus mathématiquement. Les participants sont aussi consultés pour le choix des étiquettes conceptuelles de chaque regroupement.

Les participants ne retrouvent pas nécessairement leur travail individuel dans cette représentation visuelle finale. Ils doivent donc délaisser leur propre conception pour passer à une représentation qui englobe l'ensemble du domaine conceptuel et des perspectives des participants ; ce saut de l'individu au groupe peut s'avérer ardu. L'animation de cette rencontre requiert habileté et doigté chez le modérateur et elle doit être bien planifiée.

1.3.6. Utiliser les cartes et les résultats

Comme décrit dans la section 1.2.1, la cartographie de concepts peut avoir de multiples finalités et utilités, dont l'élaboration d'un cadre conceptuel ou la définition de besoins.

2. EXEMPLE D'APPLICATION

La cartographie de concepts a été utilisée dans un projet de recherche multicentrique québécois sur les ressources résidentielles pour les adultes avec un trouble mental. La phase I du projet visait à définir les composantes clés du logement et des ressources résidentielles en santé mentale et à produire un modèle conceptuel. La complexité de l'intervention ciblée de même que l'importance de consulter de multiples parties prenantes pour la construction d'un modèle à la fois empirique et théorique faisaient de la cartographie de concepts une technique de choix (Felx, 2013).

2.1. Déroulement de la démarche

Le déroulement de la phase I du projet suit les six étapes présentées dans la section 1.3 de ce chapitre (voir la figure 23.1).

2.1.1. Préparer le processus

L'échantillonnage par choix raisonné visait à recueillir la plus grande diversité de points de vue. L'hétérogénéité des participants et l'accès limité de certains d'entre eux à Internet ont incité l'équipe de recherche à réaliser une grande partie de la collecte de données au moyen de rencontres de groupe. Les participants regroupent quatre types de parties prenantes provenant des secteurs publics, privés et communautaires provenant de cinq régions du Québec : 1) des personnes utilisatrices de services ; 2) des proches ; 3) des propriétaires, des intervenants œuvrant dans des ressources résidentielles et des intervenants responsables du suivi des résidents hébergés ; 4) des administrateurs, des gestionnaires et des intervenants responsables du suivi des ressources résidentielles. Le nombre de participants a varié selon les étapes de la démarche. Au total, pour la phase I, il y a eu 722 participations. Plusieurs participants ont contribué à plus d'une étape de la cartographie de concepts (Felx, 2013)[7].

2.1.2. Générer des idées

Treize rencontres de remue-méninges de deux heures en moyenne ont été réalisées dans quatre des cinq régions ciblées. Les participants ont été rencontrés en groupes homogènes (les types de parties prenantes ont été rencontrés séparément) de 5 à 30 personnes (n = 221). La taille des groupes a varié selon le type de parties prenantes. Les groupes composés

7. Le nombre de participants uniques est estimé à 500.

de personnes utilisatrices de services et de proches étaient habituellement plus petits (entre cinq et onze participants). Les idées étaient émises à partir de la question suivante: «Quelles sont les caractéristiques des ressources résidentielles et du logement en santé mentale (celles qu'elles ont ou celles qu'elles devraient avoir)?»

La question peut sembler enfreindre la règle de l'énoncé univoque. Toutefois, l'équipe de recherche estimait crucial de considérer l'ensemble de l'éventail des ressources résidentielles et du logement tout en permettant aux participants de spéculer sur une ressource idéale (les caractéristiques *qu'elles devraient avoir*). La lecture de l'énoncé était accompagnée d'un montage PowerPoint explicatif présentant une ressource et différentes caractéristiques. Ainsi, le terme «caractéristiques» renvoie, à la fois, à la structure du milieu (p. ex., la localisation, le nombre de membres du personnel) et à son fonctionnement (p. ex., les interventions du personnel, l'ambiance). Un prétest auprès d'un groupe d'intervenants et de gestionnaires (n = 10) a permis de s'assurer de la bonne compréhension de la question. Le prétest reproduisait le déroulement de l'étape 2.

Chaque rencontre a débuté par une brève présentation du projet de recherche. À la suite de la lecture de la question et de la visualisation du montage, cinq minutes de réflexion individuelle étaient allouées aux participants. Chacun avait à sa disposition du papier et des crayons pour écrire ses idées, si désiré. Le modérateur demandait ensuite aux participants d'émettre une idée, à tour de rôle. Au fur et à mesure, les idées étaient transcrites sur un ordinateur portable et projetées sur un écran. Les tours de table cessaient lorsque le modérateur percevait un ralentissement du rythme des idées émises. Les idées étaient alors exprimées en vrac. Après une pause de 15 minutes, les participants prenaient quelques minutes pour relire les énoncés, puis le partage d'idées reprenait, jusqu'à l'épuisement des idées (ou la fin de la rencontre). Pour clore la rencontre, le modérateur demandait aux participants s'ils étaient satisfaits de leur travail et les invitait à participer à la rencontre suivante (étape 3), qui allait se dérouler quelques mois plus tard dans leur région.

Les idées émises ont été accumulées de région en région. Une analyse partielle des idées émises était effectuée parallèlement. L'équipe de recherche a donc pu constater que très peu d'idées nouvelles avaient été exprimées par les participants de la quatrième région visitée, ce qui traduit une saturation des idées émises. Au total, 1382 idées ont été formulées (99,5% de saturation). Compte tenu du nombre d'idées recueillies, ces dernières ont été analysées à l'aide du logiciel ATLAS/ti (analyse de contenu et codage mixte). Les idées ont été regroupées en 236 codes (p. ex., *Pouvoir avoir des relations amoureuses ou sexuelles dans son milieu de vie*), 50 sous-thèmes (p. ex., *Fonctionnement*

non restrictif) et 9 thèmes (p. ex., *Règles et fonctionnement*), inspirés du cadre conceptuel du projet (Felx, 2013). Lors de la sélection des idées pour la création de la liste finale, plusieurs codes ont été exclus (p. ex., les codes rattachés aux résultats attendus de l'hébergement ou à la législation provinciale). Malgré les efforts pour réduire le nombre d'idées, la liste finale en comptait 140. Leur formulation a été très peu modifiée pour en respecter le sens et pour demeurer près du langage des participants.

2.1.3. Organiser (coter et regrouper) les idées

Un second prétest a été réalisé auprès du groupe d'intervenants et de gestionnaires qui avaient participé au premier prétest, cette fois dans l'optique d'explorer le déroulement de l'étape 3. À la suite des commentaires reçus, les deux tâches qui composent l'étape 3 ont été séparées et inversées. Dans un premier temps, les parties prenantes ont accordé une valeur numérique à chacune des 140 idées retenues au moyen de deux cotes (échelles de Likert à cinq niveaux). La première cote ciblait l'importance (1 = pas du tout importante ; 5 = très importante). La deuxième mesurait la présence ou l'absence de chacune des idées[8] dans le réseau actuel des services résidentiels en santé mentale (1 = jamais présente ; 5 = toujours présente). Par exemple, pour l'un des deux critères (importance), les participants ont répondu à la question suivante : « Selon vous, cette caractéristique possible des services résidentiels est-elle importante ? »

Au total, 41 rencontres de groupe de deux heures en moyenne ont été réalisées dans cinq régions (n = 416). Comme à l'étape précédente, la taille du groupe a varié selon le type de parties prenantes. Lors des rencontres de groupe, le modérateur lisait une à une les 140 idées. Chaque idée était simultanément projetée sur un écran. Les participants encerclaient sur leur questionnaire les deux cotes accordées à une idée (importance et présence), avant de passer à l'idée suivante.

Enfin, 73 des participants ont aussi classé les 140 idées en regroupements. Les idées ont été imprimées sur des cartons (20 × 5 centimètres). Les idées et 20 cartons vierges ont été envoyés par courrier postal aux participants, en plus de brèves consignes, d'une fiche d'identification et d'une enveloppe-réponse timbrée. Le nombre de regroupements créés par les participants a varié entre 2 et 20 (moyenne de 10 environ). Le temps requis pour classer les idées était variable. Pour certains participants, la durée totale était de plus de cinq heures.

8. Les idées représentent ici des caractéristiques structurales ou procédurales des ressources résidentielles et du logement.

2.1.4. Analyser les résultats et produire les résultats

Les données ont été analysées à l'aide du prologiciel Concept System® et du logiciel SPSS. La section 2.2 présente certaines des cartes conceptuelles et des graphiques obtenus.

2.1.5. Valider les résultats auprès des participants et interpréter les cartes

Le retour auprès des participants a pris la forme d'un groupe de discussion focalisée avec les quatre types de parties prenantes (n = 12). Les participants ont suggéré des appellations pour les différents regroupements. Certaines idées ont été déplacées vers un regroupement adjacent et les regroupements ont été fusionnés en domaines (Felx, 2013). Bien que la migration d'idées vers d'autres regroupements soit possible avec le prologiciel Concept System®, l'interface graphique permet uniquement le déplacement des frontières d'un regroupement (comparativement au déplacement d'une idée sur la carte). Pour éviter que les regroupements se chevauchent, les idées doivent être déplacées vers un regroupement adjacent.

Cette étape, d'une grande richesse pour l'équipe de recherche, a été laborieuse pour les participants. Plusieurs éléments y ont contribué : la quantité d'informations transposées sur les cartes, la complexité de la représentation, les idées de différents niveaux systémiques, l'étendue des regroupements qui ne reflète pas leur importance relative (mais bien la proximité conceptuelle à travers des idées plus ou moins dispersées). De plus, le partage par les parties prenantes de points de vue parfois divergents a nécessité de nombreuses interventions de la part du modérateur.

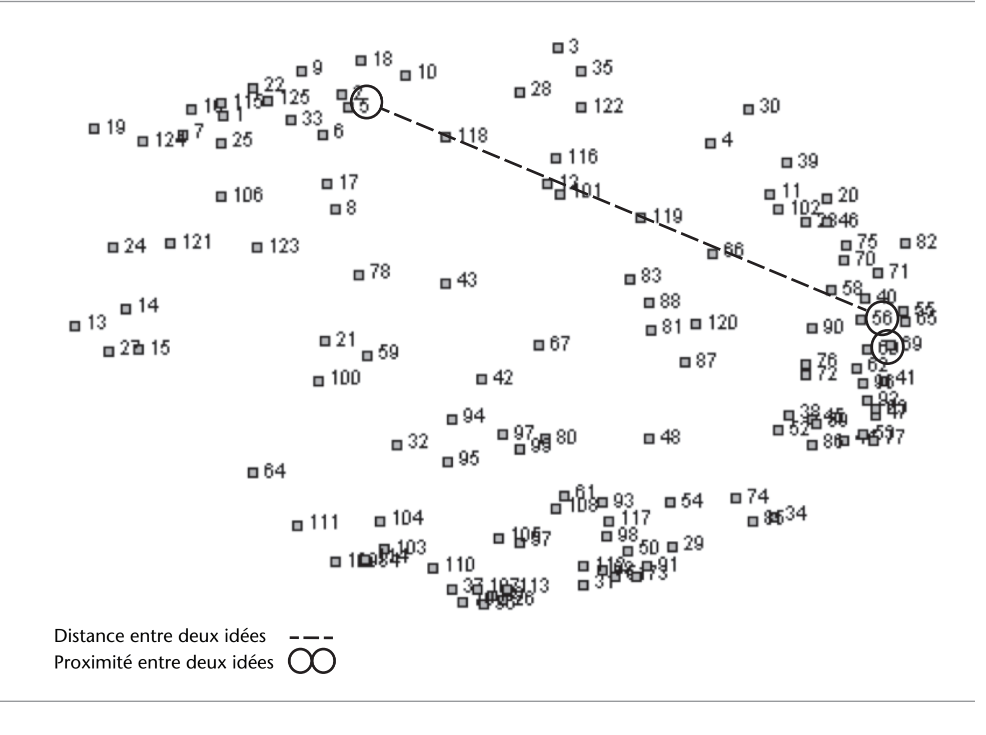
2.1.6. Utiliser les cartes conceptuelles et les autres résultats

Les résultats ont permis d'explorer les caractéristiques des ressources résidentielles en santé mentale et différents modèles conceptuels. Les données obtenues ont été utilisées pour créer un instrument de mesure qui décrit les ressources selon leurs caractéristiques structurales et procédurales. La conception de l'instrument a été réalisée en intégrant les idées et codes issus de l'étape 2 de la cartographie de concepts. Des analyses par composantes principales (voir le chapitre 22 de cet ouvrage) ont été réalisées en utilisant les résultats issus des analyses par typologie hiérarchique, ce qui a permis d'organiser et de raffiner les sections de l'instrument. La phase II du projet prévoit la poursuite de la conception de l'instrument et sa validation (Felx, 2013).

2.2. Cartes conceptuelles et autres résultats (étape 5)

La carte de la distribution géographique des idées (figure 23.3) confirme la prémisse à la base de l'échelonnage multidimensionnel. Par exemple, les idées n° 60 *Aider la personne à gérer elle-même ses médicaments* et n° 56 *Donner un accompagnement à la personne pour qu'elle puisse utiliser le transport en commun* dans la partie droite de la carte se rapportent à un même concept sous-jacent (services axés sur le développement du résident). En revanche, l'idée n° 5 *Avoir un superviseur ou un intervenant ressource qui rend visite plusieurs fois par mois aux responsables (le personnel) des ressources résidentielles* est distante des deux autres sur la carte et sur le plan conceptuel.

Figure 23.3
Carte de la distribution géographique des idées (*point map*) (*n* = 73)



Pour créer la carte des regroupements d'idées issues de la typologie hiérarchique, des solutions comprenant entre 5 et 18 regroupements ont été explorées. L'équipe de recherche a retenu la solution à 12 regroupements présentée à la figure 23.4. Dans l'ensemble, il est possible de dégager un sens pour chaque regroupement, bien que le centre de la carte compte un certain nombre d'idées hétéroclites. Le tableau 23.4 présente des exemples de deux

regroupements et des idées qui les composent. Afin de mieux comprendre la répartition des idées dans l'espace, l'équipe de recherche a créé des cartes distinctes pour les différents types de parties prenantes et a exploré les 73 classifications individuelles. L'analyse des cartes des regroupements par type de parties prenantes révèle plus de similitudes que de différences; plusieurs regroupements sont stables et ont de faibles indices de dispersion (regroupements 8, 9, 12, 10, 1, 11, 6, 3, 5). De plus, la représentation demeure près des données initiales, comme indiqué par le faible indice de stress pour un projet de cartographie de concepts (Rosas et Kane, 2012).

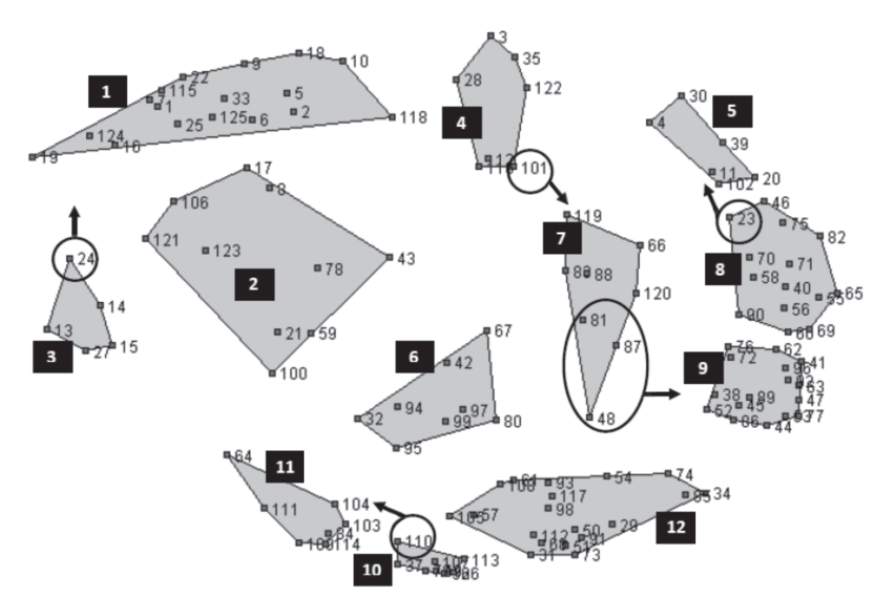
Tableau 23.4.
Exemples d'idées aléatoires comprises dans deux regroupements et leurs cotes en importance et en présence (*n* = 73)

Nom du regroupement	Indice[a]	IMP[b]	PRE[b]
8 – Services et interventions (développement) (13 idées)[c]		**4,36**	**3,39**
60. *Aider la personne à gérer elle-même ses médicaments (autogestion).*	0,04	4,11	2,99
71. *Informer la personne sur le rétablissement.*	0,21	4,43	3,17
23. *Établir/rester en contact avec la famille et l'impliquer dans le processus de soins.*	0,31	4,20	2,85
10 – Espace à soi et intimité (8 idées)		**4,32**	**3,42**
49. *Pouvoir avoir des relations amoureuses ou sexuelles dans son milieu de vie.*	0,09	4,12	2,68
110. *Être dans un milieu de vie confortable, beau, propre et qui sent bon*[d].	0,14	4,61	3,92
26. *Pouvoir participer à la décoration des lieux et choisir la couleur de sa chambre.*	0,13	4,28	3,22
113. *Avoir accès à un téléphone dans un endroit privé.*	0,04	4,46	3,87

[a] Indice de dispersion (***bridging index***) : un indice près de 0 indique peu de variation dans la façon de regrouper l'idée.
[b] Critères de cotation : importance (IMP) et présence (PRE) (1 à 5).
[c] Les valeurs en gras renvoient aux valeurs moyennes pour le regroupement.
[d] Les idées n° 23 et n° 110 ont été déplacées vers un regroupement adjacent.

La carte des regroupements d'idées a été présentée aux participants (étape 5). En examinant les regroupements un à un, il leur est apparu que sept idées devaient être déplacées dans un regroupement adjacent, car elles y étaient plus reliées conceptuellement (p. ex., idée n° 110, tableau 23.4 et figure 23.4). Le déplacement des idées a conduit à la création d'une nouvelle carte modifiée (figure 23.5).

Figure 23.4.
Carte des regroupements d'idées (*cluster map*) (n = 73)



Indice de stress de Krustal : 0,2302
Indique des idées qui ont été déplacées dans un regroupement adjacent

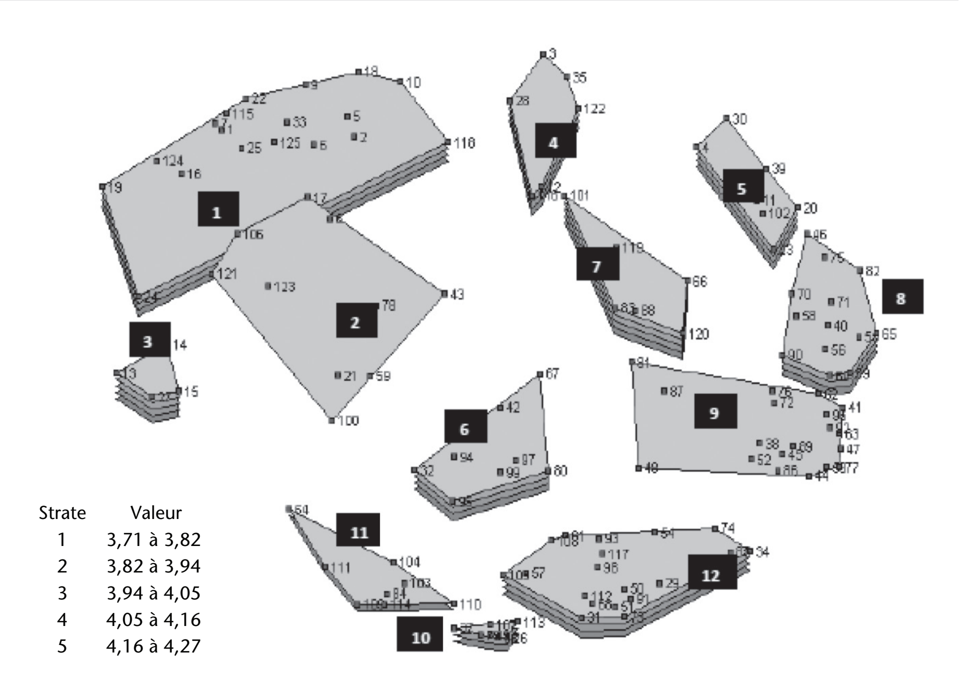
Regroupements
1 – Un système en santé et équilibré
2 – Qualité et gestion : divers
3 – Environnement physique extérieur
4 – Services résidentiels adaptés aux besoins et préférences
5 – Services et interventions (partenariat)
6 – Égalité et règlements
7 – Structure organisationnelle et personnel
8 – Services et interventions (développement)
9 – Services et interventions (quotidien)
10 – Espace à soi et intimité
11 – Environnement physique intérieur
12 – Respect, fonctionnement et atmosphère

Les résultats produits par le progiciel Concept System® permettent aussi d'explorer la valeur des idées émises individuellement ou par regroupement pour un critère donné (présence ou importance) (voir le tableau 23.4). Par exemple, l'idée n° 60 a une importance moyenne de 4,11, et celle du regroupement auquel elle se rattache est de 4,36. Deux des cartes conceptuelles intègrent les cotes moyennes accordées aux idées : la carte de la distribution des idées avec cote (*point rating map*) et la carte des regroupements d'idées avec cote (*cluster rating map*) (figure 23.5). Par exemple, la figure 23.5 permet en un coup d'œil de constater la grande importance de certains

services et interventions (regroupements 4, 5 et 8) et de l'environnement physique extérieur (regroupement 3). L'importance respective de ces différents regroupements varie entre 4,32 et 4,56, soit 4 ou 5 strates. Ces cartes peuvent être produites par type de parties prenantes ou pour l'ensemble des participants, selon l'objectif de la démarche.

Figure 23.5.

Carte des regroupements d'idées avec cote de l'importance moyenne selon les personnes utilisatrices de services (*cluster rating map*)



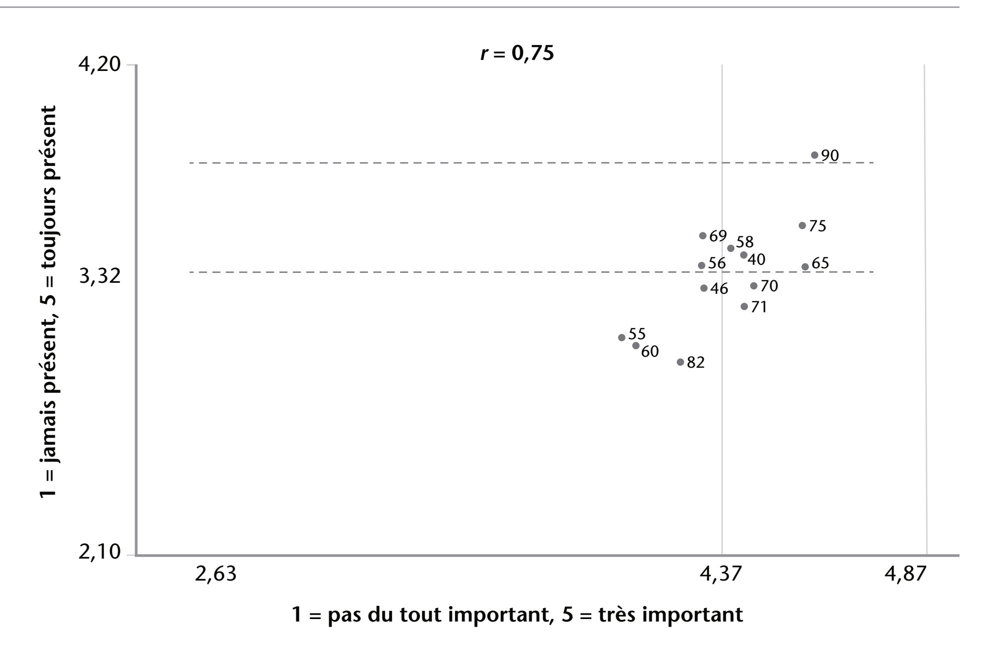
Regroupements

1 – Un système en santé et équilibré
2 – Qualité et gestion : divers
3 – Environnement physique extérieur
4 – Services résidentiels adaptés aux besoins et préférences
5 – Services et interventions (partenariat)
6 – Égalité et règlements
7 – Structure organisationnelle et personnel
8 – Services et interventions (développement)
9 – Services et interventions (quotidien)
10 – Espace à soi et intimité
11 – Environnement physique intérieur
12 – Respect, fonctionnement et atmosphère

L'utilité des graphiques additionnels varie selon la nature et les objectifs de la démarche. Par exemple, la figure 23.6 (*go-zone display*) illustre la relation entre la présence et l'importance des idées du regroupement 8,

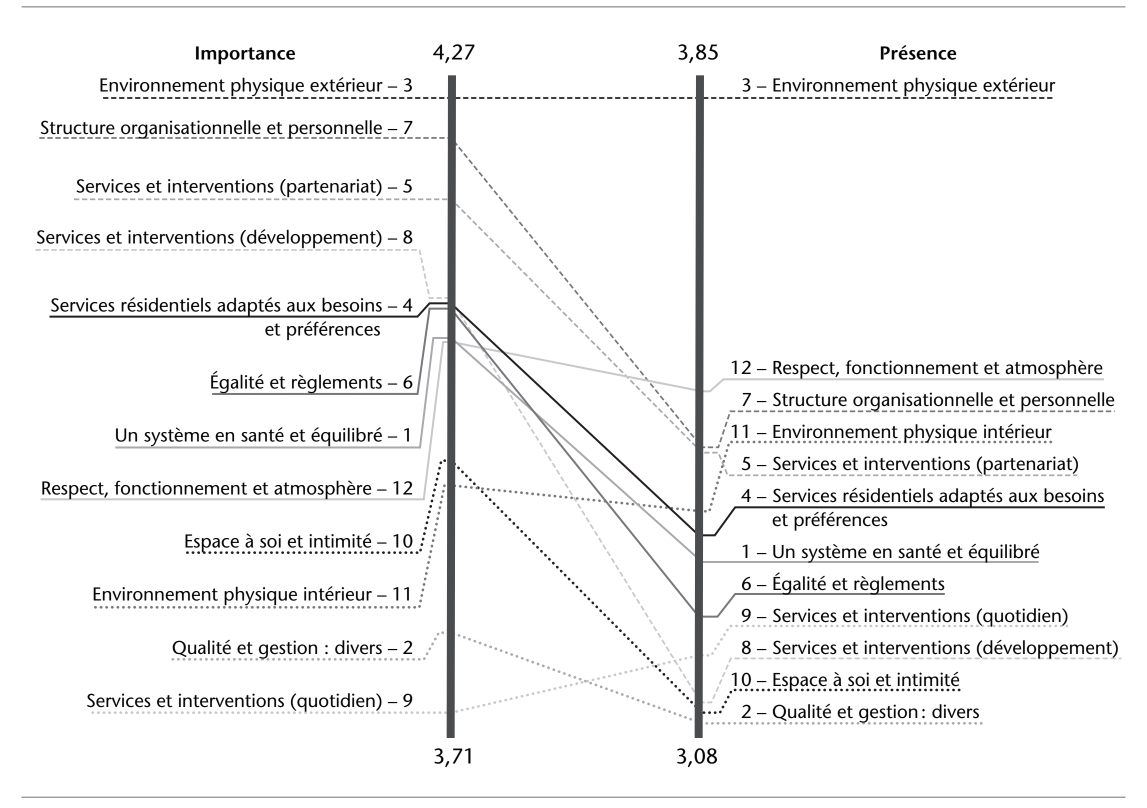
telle que perçue par l'ensemble des 73 participants. Le nuage de points, dans lequel chaque point représente une idée, montre une corrélation élevée entre la présence et l'importance des idées (r = 0,75). Il indique que quelques attributs (idées) pourraient faire l'objet d'un projet d'amélioration des services du réseau résidentiel en raison de leur faible présence et de leur importance élevée (p. ex., idées n° 71, tableau 23.4).

Figure 23.6.
Relation entre la présence et l'importance des idées du regroupement 8 (*go-zone*)



La figure 23.7 (*pattern matching display*) illustre l'écart quasi systématique entre l'importance moyenne élevée des regroupements et leur présence plus faible. Les figures 23.6 et 23.7 peuvent être produites selon certaines caractéristiques sociodémographiques (p. ex., le genre, la région), ou pour un type de parties prenantes, ou pour l'ensemble des participants.

Figure 23.7.
Relations entre la présence et l'importance attribuées aux regroupements par les utilisateurs de services (*pattern matching*)



2.3. Conseils pratiques et analyse critique de la cartographie de concepts

La cartographie de concepts a permis d'atteindre un taux de participation élevé et une mobilisation des parties prenantes au-delà des attentes de l'équipe de recherche. Les participants ont particulièrement apprécié la deuxième étape de la cartographie de concepts, lors de laquelle ils ont pu s'exprimer librement tout en visualisant simultanément le produit de leur remue-méninges. L'étonnante richesse de l'information recueillie a offert à l'équipe de recherche une connaissance approfondie des enjeux réels entourant le logement et les ressources résidentielles en santé mentale, et ce, selon divers points de vue.

L'application de la cartographie de concepts a aussi permis de constater l'investissement requis de la part de l'équipe de recherche et des participants. La technique doit être adaptée à son contexte d'utilisation, d'où l'importance du travail préparatoire et des choix de l'équipe de recherche (étape 1). Cela dit, le travail préparatoire ne met pas l'équipe de recherche à l'abri des défis. Le tableau 23.5 présente les principaux défis rencontrés lors de chacune des étapes de la démarche, ainsi que certaines des solutions retenues.

Tableau 23.5.
Défis rencontrés durant la démarche

Étapes et sous-étapes	Défis	Pistes de solutions retenues
Étape 1 – Préparer la démarche		
Formulation d'un énoncé	Complexité de l'énoncé à formuler.	Créer un montage PowerPoint et consulter des parties prenantes.
Choix des méthodes	Écart important entre les participants quant à leur degré de scolarité et leur accès au Web.	Effectuer des rencontres en face à face. Lecture des idées par le modérateur (étape 3).
Parties prenantes	Enjeux hiérarchiques et contexte politique conflictuel entre les parties prenantes.	Créer des groupes plus homogènes par type de parties prenantes. Permettre aux participants d'écrire leurs idées sur des feuilles et de les remettre au modérateur à la fin de la rencontre (étape 2).
Étape 2 – Générer les idées		
Génération des idées	Taux de participation élevé ainsi que nombre élevé d'idées émises.	Utiliser un logiciel (ATLAS/ti) pour réduire et analyser les idées (analyse de contenu).
Validité des idées	Obtention de résultats valides.	Utiliser la saturation comme critère pour la collecte de données. Réaliser une démarche qui se rapproche d'une enquête de grande échelle. Utiliser un cadre conceptuel.

(*suite*)

Tableau 23.5. (*suite*)

Étape 3 – Organiser les idées		
Organisation des idées	Nombre élevé d'idées et leur diversité.	Scinder les deux tâches (cotation et regroupement des idées).
Regroupement des idées	Tâche laborieuse pour les participants.	Réduire le nombre de participants pour cette étape. Réaliser la classification de façon individuelle (courrier postal). Encourager les participants à contacter l'équipe de recherche s'ils rencontrent des difficultés ou ont des questions.
Cotation des idées	Variabilité du temps requis pour accomplir la tâche (80 minutes à 5 heures).	Laisser les participants décider s'ils souhaitent poursuivre ou non (une deuxième rencontre peut être planifiée). Offrir du soutien aux participants durant la tâche (p. ex., relire avec un participant l'idée et les choix de réponse).
Étape 4 – Analyser les données et produire les résultats		
Analyse des données	La version utilisée du prologiciel ne permettait pas de réaliser des analyses supplémentaires (p. ex., courbes de distribution). Utilisation du coefficient de Pearson pour des données ordinales et paramétriques.	Entrer les données à la fois dans SPSS et dans Concept System®.
Analyse des données	Plusieurs éléments affectent la configuration finale de l'échelonnage multidimensionnel (Weller et Romney, 1988). La variabilité du nombre de regroupements créés doit, entre autres, être considérée.	Analyser les résultats pour l'ensemble des participants. Par la suite, exclure ceux ayant moins de quatre regroupements (n = 1) pour la configuration finale (voir Felx, 2013).
Étape 5 – Valider les résultats auprès des participants et interpréter les cartes		
Validation	Complexité de la représentation présentée. Voir section 2.1.4	Préparer le modérateur et la rencontre.
Interprétation	Juger de la qualité de la démarche et des représentations.	Examiner les biais introduits (données manquantes non aléatoires), les indices produits et la littérature existante.

CONCLUSION

Encore peu utilisée dans le domaine de la réadaptation psychosociale, la cartographie de concepts correspond à une démarche rigoureuse et structurée qui repose sur des analyses multivariées. Elle met activement à contribution les participants et l'équipe de recherche, et ce, aux différentes étapes de la démarche. Établie sur des fondements mathématiques, elle repose sur le jugement et les schémas cognitifs des participants. Elle peut décrire en une image des phénomènes complexes tout en rendant les résultats des travaux de recherche attrayants et exhaustifs. La cartographie de concepts requiert l'usage de logiciels d'analyses de données sophistiqués sans toutefois en être complètement dépendante. Pour finir, la cartographie de concepts semble pouvoir relier ce que plusieurs considèrent comme inconciliables, soit les approches qualitatives, quantitatives et participatives.

RÉFÉRENCES

ANDERBERG, M.R. (1973). *Cluster Analysis for Applications*, Duluth, Academic Press.

DAGENAIS, C., V. RIDDE, M.-C. LAURENDEAU et K. SOUFFEZ (2009). « La méthode de cartographie conceptuelle pour identifier les priorités de recherche sur le transfert des connaissances en santé des populations : quelques enjeux méthodologiques », *Revue canadienne d'évaluation de programme*, vol. 23, p. 61-80.

EVERITT, B. (1980). *Cluster Analysis*, New York, Halsted Press.

FELX, A. (2013). *Les caractéristiques structurales et procédurales des ressources résidentielles pour adultes avec un trouble mental : développement d'un instrument de mesure*, Thèse de doctorat, Montréal, Département de psychiatrie, Université de Montréal.

FOWLER, F.J. (2009). *Survey Research Methods*, 4e éd., Thousands Oaks, Sage Publications.

JACKSON, K.M. et W.M. TROCHIM (2002). « Concept mapping as an alternative approach for the analysis of open-ended survey responses », *Organizational Research Methods*, vol. 5, p. 307-336.

JAWORSKA, N. et A. CHUPETLOVSKA-ANASTASOVA (2009). « A review of multidimensional scaling (MDS) and its utility in various psychological domains », *Tutorials in Quantitative Methods for Psychology*, vol. 5, p. 1-10.

JOHNSEN, J.A., D.E. BIEGEL et R. SHAFRAN (2000). « Concept mapping in mental health : Uses and adaptations », *Evaluation and Program Planning*, vol. 23, p. 67-75.

KANE, M. et W.M.K. TROCHIM (2007). *Concept Mapping for Planning and Evaluation*, Thousand Oaks, Sage Publications.

KIKKERT, M.J., A.H. SCHENE, M.W. KOETER, D. ROBSON, A. BORN, H. HELM, M. NOSE, C. GOSS, G. THORNICROFT et R.J. GRAY (2006). « Medication adherence in schizophrenia : Exploring patients', carers' and professionals' views », *Schizophrenia Bulletin*, vol. 32, p. 786-794.

KRUSKAL, J.B. et M. WISH (1978). *Multidimensional Scaling*, Beverly Hills, Sage Publications.

MARKHAM, K.M., J.J. MINTZES et M.G. JONES (1994). « The concept map as a research and evaluation tool : Further evidence of validity », *Journal of Research in Science Teaching*, vol. 31, p. 91-101.

NOVAK, J.D. et D.B. GOWIN (1984). *Learning How to Learn*, Cambridge, Cambridge University Press.

OSBORN, A.F. (1948). *Your Creative Power*, New York, Scribner.

PINAULT, R. et C. DAVELUY (1986). *La planification en santé : concepts, méthodes, stratégies*, Paris, Éditions Nouvelles.

ROSAS, S.R. (2005). « Concept mapping as a technique for program theory development : An illustration using family support programs », *American Journal of Evaluation*, vol. 26, p. 389-401.

ROSAS, S.R. et M. KANE (2012). « Quality and rigor of the concept mapping methodology : A pooled study analysis », *Evaluation and Program Planning*, vol. 35, p. 236-245.

ROSENBERG, S. et M.P. KIM (1975). « The method of sorting as a data-gathering procedure in multivariate research », *Multivariate Behavioral Research*, vol. 10, p. 489-502.

SOUTHERN, D.M., D. YOUNG, D. DUNT, N.J. APPLEBY et R.W. BATTERHAM (2002). « Integration of primary health care services : Perceptions of Australian general practitioners, non-general practitioner health service providers and consumers at the general practice primary care interface », *Evaluation and Program Planning*, vol. 25, p. 47-59.

TROCHIM, W.M. (1989). « Concept mapping for evaluation and planning », *Evaluation and Program Planning*, vol. 12, p. 111.

TROCHIM, W. (1993). « The reliability of concept mapping », Communication présentée au congrès annuel de l'American Evaluation Association, Dallas, Texas.

TROCHIM, W., J. COOK et R. SETZE (1994). « Using concept mapping to develop a conceptual framework of staff views of a supported employment program for persons with severe mental illness », *Consulting and Clinical Psychology*, vol. 62, p. 766-775.

WELLER, S.C. et A.K. ROMNEY (1988). *Systematic Data Collection*, Thousand Oaks, Sage Publications.

CHAPITRE 24

LA CONCEPTION, LA VALIDATION, LA TRADUCTION ET L'ADAPTATION TRANSCULTURELLE D'OUTILS DE MESURE
Des exemples en santé mentale et travail

Marc Corbière
Franco Fraccaroli

FORCES

- Elles permettent l'utilisation d'outils validés dans le milieu clinique et celui de la recherche.
- Elles consistent en une approche rigoureuse qui tient compte des différences contextuelles, culturelles et linguistiques.
- Elles permettent des comparaisons internationales sur des concepts complexes.

LIMITES

- La conception d'un outil de mesure comme sa validation sont des processus exigeants et toujours perfectibles.
- La validation d'un outil de mesure se fait *in situ* et, par conséquent, elle reste contextuelle.
- La traduction et l'adaptation transculturelle exigent des chercheurs des connaissances statistiques et linguistiques avancées.

La conception d'outils de mesure vise à évaluer de nouveaux concepts ou construits, pour lesquels il n'existe pas d'outils de mesure satisfaisants dans la littérature scientifique. Le processus de conception et de validation d'un outil de mesure (p. ex., questionnaire) comprend plusieurs étapes, qui peuvent être résumées par la planification, l'opérationnalisation du concept à l'étude et sa validation. Ainsi, le chercheur qui envisage de concevoir un outil de mesure doit préalablement définir le concept à mesurer, déterminer le but de l'outil et le mode d'évaluation, et, enfin, définir la population à qui il s'adressera. La deuxième étape consiste à opérationnaliser le concept à l'étude en vue de générer le contenu de l'outil, c'est-à-dire la consigne, les items (ou énoncés) et les échelles de réponse qui le composent. La validation d'un outil de mesure est la troisième et dernière étape. Elle se traduit par l'évaluation de ses qualités psychométriques, dont les plus communes sont la validité et la fidélité. Quant à la traduction et à l'adaptation transculturelle, elles peuvent avoir lieu lorsqu'un outil de mesure est déjà existant et qu'il a fait ses preuves dans la littérature scientifique. Elles comprennent également de nombreuses étapes, dont les principales sont la traduction inversée (*back translation*) et l'atteinte de la meilleure équivalence possible (p. ex., sémantique, conceptuelle, opérationnelle) d'un contexte à un autre (p. ex., culture), lesquelles étapes sont suivies et révisées par un comité de recherche. Tout comme le processus de conception et de validation d'un outil de mesure, les qualités psychométriques de l'outil traduit et adapté au nouveau contexte nécessitent d'être évaluées. Ce chapitre est divisé en deux grandes parties. Les étapes de conception et de validation d'un outil de mesure alimenteront la première partie, alors que la seconde sera réservée à la traduction et à l'adaptation transculturelle d'un outil de mesure. Des exemples en santé mentale et travail seront aussi présentés dans le chapitre pour rendre la matière plus accessible au lecteur.

1. PLANIFICATION, OPÉRATIONNALISATION ET VALIDATION D'UN OUTIL DE MESURE

1.1. Description générale des outils de mesure : fonction, mode et nature des évaluations

L'utilisation d'outils de mesure[1] en milieu clinique et dans le cadre de projets de recherche a toujours été une activité essentielle pour une évaluation rigoureuse et précise de clients provenant d'un milieu clinique ou de participants

1. Dans le cadre de ce chapitre, les outils de mesure peuvent faire référence aux tests (ex. tests cognitifs), aux questionnaires administrés par un intervenant ou à ceux qui sont autorapportés, ainsi qu'aux entretiens directifs ou semi-directifs.

consentants à la recherche. Trois fonctions sous-jacentes à l'utilisation des outils de mesure sont à considérer : la fonction discriminative, la fonction évaluative et la fonction prédictive. Alors que la fonction discriminative vise à classer et à ordonner des caractéristiques personnelles individuelles (p. ex., comportements), la fonction évaluative cherche à documenter les éventuels changements chez la personne évaluée. La fonction prédictive, quant à elle, tente de prédire un comportement, une attitude chez l'individu et ainsi préciser son évolution dans le temps (Corbière, 2011). Ces trois fonctions de l'évaluation sont non seulement reliées aux types de validité que nous aborderons plus tard dans ce chapitre, mais peuvent aussi à leur tour s'actualiser par divers modes d'évaluation[2], dont les plus courants sont les suivants : les instruments de mesure (p. ex., un thermomètre) ; l'observation (p. ex., un superviseur observera derrière une glace sans tain les comportements d'un étudiant en session de psychothérapie) ; les questionnaires autorapportés (un participant d'une recherche remplira un questionnaire pour évaluer son sentiment d'efficacité pour réaliser une tâche) ; l'entretien, qui peut être directif, semi-directif ou non directif.

Ces différents modes d'évaluation ne sont pas équivalents et présentent chacun leurs forces et leurs limites. Dans le cadre d'une évaluation quantitative, les questionnaires autorapportés présentent toutefois de nombreux avantages lorsque comparés avec les autres modes d'évaluation, comme un moindre coût, une passation rapide notamment pour de grands échantillons, une faible influence de la part de l'évaluateur sur les réponses du répondant et une approche systématique pour l'ensemble des participants à l'étude (Corbière, 2011). D'ailleurs, dans le cadre de ce chapitre, la présentation et l'exemple de la conception et de la validation d'outils de mesure passeront essentiellement par les questionnaires autorapportés.

Avant de poursuivre, il nous semble important d'ajouter quelques informations relatives à la nature des modes d'évaluation, qui peut se résumer par une présence en face à face, par téléphone, par courrier et en ligne. En fait, il n'y a pas de mode d'évaluation idéal, cela dépend du type de l'étude (p. ex., degré de complexité, durée de la passation) et de la population cible (p. ex., population générale, taille, facilité d'accès) et des ressources financières et humaines dont l'équipe de recherche dispose (p. ex., subventions, professionnels de recherche). Les avantages et les inconvénients reliés à la nature des modes d'évaluation sont rapportés dans le tableau 24.1 (Streiner et Norman, 2008).

2. La nature de ces modes d'évaluation peut aussi être différente (p. ex., par téléphone, en face à face, en ligne).

Tableau 24.1.
Avantages et inconvénients des différents modes d'évaluation

En face à face	
Avantages	*Inconvénients*
▪ La personne est bien identifiée. ▪ Le risque d'omission de questions est très faible. ▪ Il est possible de répondre aux incompréhensions du participant. ▪ Certaines personnes ayant des difficultés pour la lecture (p. ex., langue seconde) peuvent entendre les questions plutôt que de les lire. Au besoin, les questions peuvent être reformulées (mais attention, car il y a un risque de changer le sens des questions). ▪ Les personnes peuvent répondre de vive voix plutôt que par écrit.	▪ Les coûts associés au déplacement de l'intervieweur (argent et temps). ▪ Le participant pourrait ne pas se présenter au rendez-vous fixé. ▪ Les intervieweurs doivent être formés et parfois doivent être bilingues (anglais-français). ▪ Les caractéristiques des intervieweurs (p. ex., genre, âge, attitudes) peuvent avoir un effet sur les réponses des participants à l'étude. ▪ Si un enregistrement est prévu, il y a aussi des coûts de transcription à prévoir. ▪ Si l'entretien est long, le participant peut se sentir obligé de poursuivre.
Par téléphone	
Avantages	*Inconvénients*
▪ Il y a un gain de temps et d'argent, car il n'y pas de déplacements (c'est moitié prix par rapport aux entretiens en face à face). ▪ Il y a la possibilité de cibler les régions grâce à l'indicatif téléphonique. ▪ Il y a la possibilité de faire les appels téléphoniques à partir de différents lieux.	▪ Les personnes ne sont pas toujours disponibles lors des appels. ▪ Les personnes peuvent trier les appels (numéro inconnu qui est affiché sur leur téléphone). ▪ Il est parfois difficile de vérifier l'identité de la personne. ▪ Il peut s'avérer complexe d'expliquer au téléphone les divers intervalles d'une échelle de réponse. ▪ Il y a un risque d'acquiescer chez le répondant, car les questions sont trop complexes à comprendre au téléphone (pour pallier cette difficulté, il est possible d'envoyer le questionnaire à l'avance pour s'assurer de gagner du temps et de préparer correctement l'entretien). ▪ Il est parfois nécessaire de faire des appels répétitifs pour conduire l'entretien (entre trois et six essais).

(*suite*)

Tableau 24.1. (*suite*)

Par courrier	
Avantages	*Inconvénients*
▪ Cela est moins coûteux que par téléphone. ▪ Il y a un niveau de désirabilité sociale moins grand, car il n'y a pas la présence d'intervieweurs.	▪ Lors du retour des questionnaires, il est parfois difficile de savoir qui n'a pas répondu, notamment lorsque les réponses sont anonymes (possibilité de faire un second envoi à tous avec le risque d'avoir des doublons). ▪ Les participants peuvent ne pas répondre à toutes les questions (5 à 10 % des réponses manquantes). ▪ L'ordre des questions n'est pas nécessairement suivi. ▪ Le délai de réponse peut être plus long que par téléphone ou en face à face.
En ligne	
Avantages	*Inconvénients*
▪ Il n'y a pas besoin d'intervieweurs. ▪ Un repérage systématique des erreurs est possible. ▪ Il y a la possibilité de rapporter des éléments qui ne sont pas désirés socialement. ▪ Il y a la présence d'un support visuel en comparaison avec un entretien par téléphone. ▪ Si le participant répond à certains critères, il n'est pas obligé de poursuivre (renvoi systématique à certaines sections). ▪ L'enquête peut être réalisée en quelques jours, alors que les autres méthodes présentées ci-dessus peuvent prendre plusieurs semaines. ▪ Le coût est moins élevé qu'en entretien en face à face. ▪ Le respect de l'anonymat.	▪ Il peut y avoir une appréhension à répondre en ligne. ▪ Il peut y avoir une difficulté à revenir à la même place dans le questionnaire s'il y a une interruption (p. ex., pause). ▪ Il peut y avoir du découragement de la part du participant s'il y a un trop grand nombre de clics pour arriver au questionnaire. ▪ Il y a un risque de frais de ligne Internet pour le participant. ▪ Le message d'invitation à participer à l'enquête peut être court-circuité par des filtres Internet. ▪ Il y a un taux de réponse plus bas que celui obtenu par courrier. ▪ Il faut avoir la maîtrise de certains logiciels pour concevoir ce type d'enquête en ligne.

Source : Inspiré de Steiner et Norman, 2008.

1.2. Étapes de conception et de validation des outils de mesure

Plusieurs auteurs se sont intéressés à la conception et à la validation d'outils de mesure en proposant plusieurs étapes à suivre, dont les plus communes sont celles relatives à la planification, à l'opérationnalisation et à la validation (Benson et Clark, 1982 ; DeVellis, 2011 ; Netemeyer, Bearden et Sharma,

2003 ; Streiner et Norman, 2008). Ces trois grandes étapes seront présentées dans les prochains paragraphes pour mieux les appréhender et sont résumées dans le tableau 24.2.

Tableau 24.2.
Étapes de conception et de validation d'un outil de mesure

Étapes	Indicateurs des étapes pour éviter les écueils
Planification	▪ Formuler l'objectif de l'outil de mesure et donner une première définition du concept à l'étude. ▪ Faire une recension des écrits scientifiques pour : ▪ s'assurer que des outils de mesure n'existent pas déjà pour évaluer le concept ciblé ; ▪ prendre connaissance des théories dans lesquelles le concept s'inscrit ; ▪ bien déterminer les attributs et dimensions du concept et ainsi en dégager une définition claire et opérationnelle. ▪ Si la recension des écrits reste lacunaire quant au concept ciblé, alors : ▪ faire appel à des experts pour mieux appréhender le concept ; ▪ utiliser des méthodes qualitatives appropriées auprès des groupes utilisateurs du futur outil de mesure pour mieux définir le concept. ▪ Formuler une nouvelle définition du concept en fonction des étapes précédentes en spécifiant quelles dimensions ou facettes du concept seront opérationnalisées.
Opération-nalisation	▪ Pour chaque dimension conceptuelle retenue au préalable, la tâche est de relier un certain nombre d'items à cette première. ▪ Générer les items : ▪ formuler des items clairs et précis ; ▪ tenir compte des particularités de la population cible ; ▪ formuler un nombre d'items suffisant (un minimum de cinq items) pour chaque dimension. ▪ Choisir les items : ▪ faire attention de ne pas écarter des items trop rapidement, car considérés comme boiteux (préférable de les retravailler plutôt que d'en créer de nouveaux) ; ▪ avoir recours à un groupe d'experts pour obtenir leur opinion ; ▪ avoir recours à des analyses statistiques plus complexes (p. ex., analyse factorielle). ▪ Choisir l'échelle de réponse : ▪ prendre connaissance des divers types d'échelles possibles ; ▪ vérifier que l'échelle de réponse est appropriée pour coter chaque item (p. ex., Likert).

(*suite*)

Tableau 24.2. (*suite*)

Validation	▪ Prétester les items auprès de la population cible (étude pilote). ▪ Évaluer les différents types de validité : apparente, de contenu, de construit (c.-à-d. par analyses factorielles, convergente, discriminante, par groupes extrêmes), de critère (concomitante et prédictive). ▪ Évaluer les différents types de fidélité : cohérence interne, fidélité test-retest, fidélité interjuges. ▪ Évaluer la sensibilité au changement et la réactivité clinique (*responsiveness*). ▪ Adopter une séquence logique (non exhaustive) pour la validation d'outils mesure : 1. validité apparente (la beauté du regard !) ; 2. validité de contenu (analyses descriptives) ; 3. validité de construit par analyses factorielles (analyses factorielles) ; 4. cohérence interne (calcul du coefficient alpha de Cronbach) ; 5. validités convergente et divergente (corrélations) ; 6. validité par groupes extrêmes (comparaisons de moyennes) et validité concomitante (corrélations) ; 7. fidélité test-retest (corrélations) et validité prédictive (analyses de régression) ; 8. sensibilité au changement et réactivité clinique (comparaisons de moyennes) ; 9. établissement de valeurs de référence ou normes (analyses descriptives).

1.2.1. Planification

L'étape de la planification est la plus importante et probablement la plus ardue à réaliser. Par conséquent, elle ne doit pas être négligée. À cette étape, le chercheur qui envisage la conception d'un outil de mesure se doit de clairement définir le concept qui est à mesurer, d'établir dans quelles théories il s'inscrit, de déterminer le but de l'outil et son mode d'évaluation, et enfin de définir la population à qui il s'adresse (c.-à-d. le participant ciblé). Aussi, le chercheur se posera les questions suivantes : à quel degré l'outil de mesure doit-il être spécifique ? Souhaite-t-on évaluer une ou plusieurs dimensions du concept à l'étude ? (DeVellis, 2011). Pour définir le concept à la base d'un outil, les auteurs peuvent procéder à une recension des écrits sur le sujet ou s'appuyer sur un cadre théorique pour comprendre comment ce concept s'articule à d'autres concepts (c.-à-d. relations d'influence) (Clark et Watson, 1995). Parmi les méthodes qualitatives utiles à cette étape, on compte par exemple l'analyse de concept (voir le chapitre 6 de cet ouvrage) et la revue systématique (voir le chapitre 7 de cet ouvrage). Le cadre théorique dans lequel s'inscrit le concept est sans conteste d'une grande utilité pour une meilleure connaissance et définition du concept (DeVellis, 2011), mais aussi pour bien le circonscrire en écartant ce qu'il n'est pas (Clark et Watson, 1995). Worthington et Whittaker (2006) mentionnent que cela exige du chercheur de bien définir chaque attribut du concept à l'étude et d'acquérir ainsi une compréhension fine et exhaustive des causes et

des conséquences relatives à ce concept. Cette démarche sera utile non seulement pour soutenir les validités de contenu et de construit (Netemeyer *et al.*, 2003), mais aussi pour ne pas trop restreindre la cible de la recension des écrits dès le début du processus (Clark et Watson, 1995). À la suite de la recension des écrits, si un outil de mesure est repéré, il est alors important de se poser les questions suivantes : est-ce que la conception d'un nouvel outil de mesure est requise, et si oui, pourquoi ? À quelles fins ? Et à quel coût ? Contandriopoulos *et al.* (2005) précisent à cet effet qu'il existe au moins trois inconvénients inhérents à la conception d'un nouvel outil de mesure, notamment une demande de temps et des ressources, un éventuel retard pour commencer l'étude et la nécessité de préparer une étude pilote pour prétester la première version de l'outil de mesure. Streiner et Norman (2008) mentionnent que si un nouvel outil est élaboré, c'est que plusieurs raisons appuient cette décision. Les questions qui suivent peuvent permettre au chercheur d'étayer sa décision : l'outil déjà existant est-il coûteux, invasif, dangereux, trop long à administrer, trop général ? Lors de la recension des écrits, il est en effet possible de constater qu'un ou plusieurs outils de mesure sont disponibles pour évaluer le concept à l'étude, avec toutefois des qualités psychométriques insatisfaisantes (voir plus bas pour ces notions), une clientèle ciblée qui n'est pas spécifique ou encore un format inadéquat (p. ex., trop long ou trop court). L'ensemble de ces éléments amène le chercheur à concevoir un nouvel outil pour répondre à l'objectif d'évaluation qu'il s'est fixé. En bref, dans la mesure où aucun outil n'est disponible dans la littérature scientifique ou encore parce que les outils ne sont pas satisfaisants, le chercheur pourra alors décider de concevoir un nouvel outil de mesure ou questionnaire autorapporté, tout en sachant qu'il aura besoin de temps et de ressources supplémentaires.

En parallèle, il s'avère important de déterminer à quelles fins sera utilisé l'outil de mesure. Un outil de mesure conçu pour la recherche est généralement plus court et moins détaillé qu'un outil conçu pour la clinique, quoique ces différences s'estompent graduellement quand un outil d'abord conçu pour la recherche est de plus en plus utilisé en clinique. De plus, d'autres précautions sont à prendre, notamment en répondant à la question suivante : l'outil sera-t-il générique (p. ex., évaluation de l'estime de soi générale) ou spécifique (p. ex., évaluation de l'estime de soi en tant que travailleur) ? Le chercheur doit aussi déterminer pour quelle clientèle il souhaite concevoir l'outil de mesure, tout en ayant au préalable bien circonscrit les caractéristiques de la clientèle ciblée. Par exemple, la conception d'un questionnaire dont la population cible est un groupe de personnes ayant un faible niveau de scolarité exigera du chercheur l'utilisation d'un vocabulaire et d'un format adaptés à leur niveau de compréhension et de connaissances.

Lorsqu'il existe peu d'écrits sur le concept ciblé, le chercheur peut faire appel à des personnes clés (experts, cliniciens, clients, etc.) afin d'élaborer une définition claire du concept et sur laquelle le chercheur s'appuiera pour formuler le contenu de l'outil de mesure. Il aura aussi le souci de rédiger sa propre compréhension du concept avant même de l'opérationnaliser (DeVellis, 2011 ; voir *epoché* dans le chapitre 2 de cet ouvrage). Dans ce cas de situation, des méthodes qualitatives (qui sont non exhaustives) sont également très pertinentes, telles que l'analyse descriptive interprétative (voir le chapitre 1 de cet ouvrage), l'analyse ethnographique (voir le chapitre 3 de cet ouvrage) ou l'étude de cas (voir le chapitre 4 de cet ouvrage). Par ailleurs, la consultation d'experts et de la théorie dans laquelle peut s'inscrire un concept n'est pas suffisante, l'évaluation du contexte et de la population dans et pour laquelle le chercheur souhaite évaluer le concept sera une condition *sine qua non* de la validité et de la fidélité de l'outil de mesure au sens large. À l'issue de l'ensemble des étapes présentées ci-dessus, le chercheur sera amené à revoir la définition du concept à l'étude qu'il aura au préalable rédigée, en précisant les attributs et dimensions qu'il souhaitera opérationnaliser.

1.2.2. Opérationnalisation

La deuxième grande étape est celle de l'opérationnalisation du concept, dont le but est de générer des items (énoncés, phrases) afin de bien représenter le contenu de l'outil. Dans le cadre d'un questionnaire autorapporté, les items et les échelles de réponse qui le composent doivent faire référence aux dimensions conceptuelles du concept plus large. Par exemple, si le chercheur souhaite évaluer les facteurs significatifs du retour au travail de personnes qui présentent un trouble mental, en s'appuyant sur les résultats de la première étape (c.-à-d. la planification), il pourra considérer plusieurs sous-dimensions conceptuelles telles que les facteurs reliés à la santé (p. ex., symptômes) et ceux reliés à l'organisation de son travail (p. ex., ressources disponibles pour effectuer son travail) ou encore ceux reliés à son équipe de travail (p. ex., climat organisationnel). L'étape de l'opérationnalisation peut se diviser en trois sous-étapes : générer des items, choisir des items et choisir les échelles de réponse appropriées.

Générer des items

Les items peuvent être élaborés à partir : 1) de la théorie et de la recension des écrits ; 2) d'une révision des items d'outils déjà existants qui mesurent le même concept ; et 3) de l'opinion des experts. Quel que soit l'objectif du chercheur, il est préconisé de générer des items qui ciblent l'ensemble

des dimensions du concept pour ne rien écarter de façon hâtive (Clark et Watson, 1995). Comme vu plus haut, la littérature scientifique est probablement le meilleur point de départ pour élaborer les items d'un questionnaire autorapporté. Elle peut fournir assez d'informations au chercheur pour lui permettre en retour de déterminer les dimensions ou attributs du concept à l'étude et de documenter chacune d'elles adéquatement (Netemeyer *et al.*, 2003; Worthington et Whittaker, 2006). Il est aussi possible pour les concepteurs du nouvel outil de s'inspirer du contenu d'outils jugés insatisfaisants ou provenant d'autres contextes, en reprenant les items qui pourraient être pertinents et donc retenus.

Par exemple, dans l'étude de Corbière, Brouwers, Lanctôt et van Weeghel (sous presse-a), l'objectif était de concevoir un outil sur les compétences des conseillers en emploi spécialisés qui œuvrent dans des programmes de soutien à l'emploi. Pour ce faire, les chercheurs se sont appuyés, d'une part, sur des outils déjà existants, notamment des questionnaires dont la population cible étaient des professionnels de la santé et du travail (psychologues, ergothérapeutes, coordonnateur de retour au travail, conseillers en emploi, etc.) et, d'autre part, sur des recherches de nature qualitative qui renvoyaient aux compétences des conseillers en emploi spécialisés pour, enfin, consulter l'avis des experts (c.-à-d. les conseillers en emploi spécialisés). Les items reflètent par ailleurs les diverses dimensions du concept à l'étude et du même coup permettent une meilleure validité de contenu de l'outil (Netemeyer *et al.*, 2003). Enfin, lorsque le chercheur est prêt à formuler des items, il ne doit pas se censurer, en considérant toutefois les éléments suivants: 1) la clarté et la spécificité, 2) la redondance, 3) la formulation négative, et 4) la double négation et le double sens, ainsi que 5) les items qui dirigent la réponse sans équivoque ou qui sont à forte connotation émotive (Clark et Watson, 1995; DeVellis, 2011; Netemeyer *et al.*, 2003; Streiner et Norman, 2008; Worthington et Whittaker, 2006). Dans les prochains paragraphes, ces éléments seront expliqués plus en détail.

L'aspect premier de la formulation des items est certainement leur *clarté*, grâce à l'utilisation d'un langage commun, concis, simple et précis, qui évite les ambiguïtés, le double sens, les jargons (p. ex., termes utilisés dans le milieu médical), les expressions régionales et le verbiage. Il est aussi recommandé de formuler une seule idée par item, mais par contre d'éviter de formuler une idée à laquelle presque tout le monde adhère. En effet, dans ce dernier cas, l'item sera de peu d'utilité pour l'évaluation finale (c.-à-d. une faible variabilité dans les réponses). Il est à noter que le degré de précision des items dépendra aussi de la fonction que l'on souhaite leur octroyer, c'est-à-dire l'évaluation d'un concept général ou de comportements spécifiques (DeVellis, 2011). Cette *spécificité* dépendra en majeure partie du

nombre de dimensions que l'outil comportera ou encore de l'objectif que ce sera fixé le chercheur, comme développer une dimension particulière d'un concept (p. ex., les symptômes cognitifs de la dépression) ou de son aspect plus global (p. ex., les symptômes généraux de la dépression). Pour la formulation des items, il est généralement recommandé de ne pas utiliser plus de 15 à 16 mots[3] par question ou énoncé tout en utilisant un vocabulaire accessible à tous. En fait, il est en général suggéré de viser un niveau de lecture d'une personne qui a terminé son secondaire. Streiner et Norman (2008) suggèrent d'éliminer des mots pour vérifier si le sens de la phrase change et ainsi tenter de rechercher la parcimonie du contenu de l'item. Plus la phrase est longue et plus on perd en fidélité (Streiner et Norman, 2008).

DeVellis (2011) mentionne qu'au début de cette étape, la *redondance* du nombre d'items peut être désirée afin de bien comprendre le phénomène dans ces moindres aspects. Toutefois, la redondance des items est à proscrire lorsqu'il s'agit de changer un seul mot dans un énoncé, car cela n'apportera que peu d'information et créera aussi des erreurs de mesure (c.-à-d. une trop grande similarité). Par contre, si l'item est formulé de façon bien différente pour évaluer un même attribut conceptuel, il peut s'avérer dans certains cas utile pour tester la cohérence du répondant dans ses réponses. Aussi, le chercheur pourra retenir parmi l'ensemble des items redondants celui qui présente la meilleure qualité sur le plan de la formulation (DeVellis, 2011). Au final, il importe de savoir que la redondance des items peut gonfler le coefficient de cohérence interne de la dimension conceptuelle à l'étude. À titre d'information, lors du calcul du coefficient de cohérence interne, ce dernier ne fait pas la distinction entre ce qui est redondant et ce qui est spécifique pour l'échelle de mesure ou sa dimension (DeVellis, 2011), d'où l'importance d'éviter la redondance accrue dans la formulation des items.

La *formulation d'items à la négative* est une autre méthode qui permet de contrôler la congruence des réponses de la personne, mais aussi pour contrecarrer la tendance du participant à acquiescer à des items qui sont tous formulés à la positive. Toutefois, les items formulés à la négative peuvent créer une certaine confusion chez le participant, sans compter les erreurs lorsqu'il est temps de calculer les scores appartenant aux échelles. Aussi, le répondant peut coter un item différemment selon qu'il est formulé à la positive ou à la négative. Streiner et Norman (2008) mentionnent à cet effet que les personnes ont tendance à approuver un item formulé à la négative plutôt que de rejeter un item formulé à la positive. Ils précisent aussi qu'inverser la polarité des items (renverser le score) ne renverse pas nécessairement le

3. Ce nombre de mots est valable pour la langue anglaise. La langue française requiert toutefois environ 20 % de mots de plus pour exprimer la même idée qu'en anglais. On peut donc penser que pour la langue française, un maximum de 20 mots par phrase ou item serait requis.

sens, sous-entendant ainsi que les personnes ne cotent pas leurs réponses de la même façon lorsque l'item est formulé à la positive ou à la négative. Par exemple, pour une personne qui répond oui à l'item *Je suis satisfait de mon travail* et qui répond non à l'item *Je ne suis pas satisfait de mon travail*, la deuxième réponse reflète l'absence d'un aspect négatif, ce qui n'est pas équivalent à la première réponse. Enfin, une échelle qui comprend à la fois des items formulés à la positive et à la négative présentera un coefficient alpha de Cronbach plus faible que si elle comprend uniquement des items formulés à la positive. Autrement dit, le chercheur gardera toujours à l'esprit qu'il est préférable de ne pas utiliser d'items formulés à la négative, car il y a plus d'inconvénients que d'avantages (DeVellis, 2011) ou encore, il devra prendre soin d'évaluer à l'avance les éventuelles conséquences de ce choix.

Par ailleurs, il est important d'éviter d'utiliser les *doubles négations* dans la formulation des énoncés (p. ex., *Je ne suis pas en mesure de ne pas demander de l'aide au besoin*), les items à *double contenu* (p. ex., *J'ai de la difficulté à réaliser des activités familiales et professionnelles*) et les mots ou phrases qui dirigent la réponse *sans équivoque* (p. ex., *Je participe à un programme de réinsertion au travail pour obtenir un emploi*) ou encore qui sont à *forte connotation émotive* (p. ex., *J'adore travailler avec autrui*). En effet, ces dernières formulations d'items sont à proscrire, car elles créent de la confusion chez le répondant ou encore affectent ses réponses, en créant des biais.

Lorsque la littérature scientifique est lacunaire, voire inexistante à propos du concept à l'étude, une des méthodes les plus souvent utilisées pour générer les items d'un outil de mesure est l'utilisation de l'opinion d'experts. Le chercheur qui emploie cette méthode doit expliquer comment il a sélectionné les experts qui ont participé à la conception de l'outil et quelles méthodes il a employées pour obtenir les informations nécessaires auprès de ces experts. Les méthodes et techniques qualitatives les plus fréquemment utilisées pour animer des groupes d'experts sont, par exemple, le groupe de discussion focalisée (voir le chapitre 12 de cet ouvrage), la méthode Delphi (voir le chapitre 13 de cet ouvrage) et la technique TRIAGE (voir le chapitre 14 de cet ouvrage). À cette étape de la conceptualisation, il y a en général beaucoup plus d'items qui sont développés pour obtenir un bassin d'items exhaustif plutôt que trop restreint (Netemeyer *et al.*, 2003). L'opérationnalisation des énoncés à partir des comptes rendus intégraux (*verbatim*) des répondants à un groupe de discussion ou d'une technique TRIAGE est probablement l'une des étapes les plus complexes, car elle demande de la part de l'équipe de recherche de prendre en considération à la fois le contenu d'une idée évoquée par un ou plusieurs participants, mais aussi la forme particulière de cette même idée (syntaxe, vocabulaire) pour au

bout du compte produire un ou plusieurs items qui soient riches de sens et de forme pour le futur participant. On comprend alors que cette étape exige du chercheur un haut niveau de rigueur et de créativité (DeVellis, 2011).

Toujours à cette étape-ci, la règle est d'adopter une approche inclusive ou plus large que trop étroite. La question soulevée par Netemeyer *et al.* (2003) est celle que les chercheurs se posent lorsque l'étape de générer les items a débuté : combien d'items cela prend-il pour mesurer le concept à l'étude ? En fait, la réponse dépendra en grande partie du nombre de dimensions que le concept comprend. Il est fortement suggéré d'utiliser plusieurs items pour évaluer chaque dimension du concept à l'étude (DeVellis, 2011 ; Netemeyer *et al.*, 2003). Un nombre de cinq à dix items par dimension est recommandé non seulement pour bien représenter cette dernière, mais aussi pour prévoir dans la version finale de l'outil un nombre minimal de trois items par dimension (minimum requis pour le calcul du coefficient de la cohérence interne). On comprend alors qu'un bassin initial de plusieurs dizaines, voire de centaines d'items peut être planifié lorsque le concept à l'étude est complexe et qu'il inclut de nombreuses dimensions (Corbière *et al.*, sous presse-a). Si le chercheur sait à l'avance combien d'items seront compris dans l'outil de mesure final, DeVellis (2011) suggère dans la mesure du possible d'en créer trois à quatre fois plus pour prévoir la perte ultérieure d'items. Clark et Watson (1995) mentionnent que si seulement deux ou trois items représentent une facette ou une dimension du concept, la version finale de l'outil sera de beaucoup réduite quant à sa représentation conceptuelle. Enfin, il peut s'avérer utile de créer un tableau pour préciser quel item appartient à quelle dimension (pour une illustration, voir Benson et Clark, 1982, p. 794).

Choisir les items

Après avoir généré des items pour concevoir l'outil de mesure, il s'agira de faire un choix parmi eux. Il est fortement recommandé que le choix des items ne soit pas fait seul (p. ex., le chercheur) mais bien en équipe. L'équipe de recherche peut comprendre à la fois des chercheurs, des cliniciens et des répondants au questionnaire, pour ne citer que quelques acteurs essentiels. La composition de l'équipe dépendra aussi du concept à l'étude, du milieu dans lequel l'outil sera utilisé, auprès de quelle population et à quelles fins. Le chercheur ou l'équipe sélectionnera assez d'items (à partir du bassin d'items généré à l'étape précédente) pour couvrir l'ensemble des dimensions du concept. Une nouvelle fois, le choix des items doit tenir compte de la fonction de l'outil, de la population cible et du contexte de l'évaluation (p. ex., clinique ou recherche). Lorsqu'il s'agit d'un questionnaire autorapporté, il est important de s'assurer de la clarté des questions

qui seront posées aux personnes qui devront le remplir et de tenir compte du niveau de lecture, qui peut varier en fonction du niveau de scolarité de la population cible.

Pour choisir les items qui seront inclus dans l'outil de mesure, le chercheur et son équipe pourront à nouveau avoir recours à l'opinion d'experts ou réaliser un prétest auprès d'un groupe de personnes faisant partie de la population cible. Comme Netemeyer *et al.* (2003) le suggèrent, les experts peuvent être invités à évaluer le degré auquel l'item est représentatif ou non du concept à l'étude (p. ex., évaluation en trois points), et le chercheur pourra retenir les items qui présentent une moyenne satisfaisante (seuil à déterminer par l'équipe) ou ceux ayant les scores les plus élevés (voir le chapitre 13 de cet ouvrage). Plusieurs auteurs recommandent de demander aux experts ce qui a été oublié pour avoir une représentation plus exhaustive du concept (DeVellis, 2011; Worthington et Whittaker, 2006). Worthington et Whittaker (2006) rappellent toutefois que la durée de passation d'un questionnaire ne devrait pas être de plus de 50 minutes et idéalement, être inférieure à 30 minutes.

D'autres types d'analyses statistiques peuvent être réalisées ultérieurement avec un nombre plus conséquent de répondants (>100) afin de vérifier s'il y a une corrélation significative entre les items ainsi qu'une redondance possible. À partir de ces résultats, les items très fortement intercorrélés peuvent être retranchés afin d'évaluer l'essence du concept à l'étude et ainsi assurer la meilleure parcimonie possible (voir le chapitre 22 de cet ouvrage). Streiner et Norman (2008) stipulent que les items qui sont corrélés entre 0,20 et 0,80 devraient être retenus, et Clark et Watson (1995) précisent qu'une corrélation moyenne interitems devrait se situer entre 0,15 et 0,50. Les calculs d'une matrice de corrélation interitems, du coefficient alpha de Cronbach ainsi que l'analyse factorielle sont souvent utilisés pour refléter la redondance possible des énoncés appartenant à une dimension conceptuelle.

Choix des échelles de réponse pour les items retenus

Une fois que les items ont été conçus[4], les auteurs devront choisir des échelles de réponse qui permettront de donner une catégorie, un rang ou un nombre à ce qui sera mesuré par les réponses aux items (Netemeyer *et al.*, 2003). C'est la nature de l'item qui détermine le type d'échelle qui devra être utilisée, mais aussi le cadre théorique dans lequel ce concept s'inscrit.

4. Cette étape peut aussi être conduite en même temps que la conception des items, pour ainsi permettre la formulation des énoncés en ayant déjà en tête le type d'échelle de réponse qui sera utilisé. Cette stratégie peut s'avérer d'ailleurs très efficace (DeVellis, 2011).

Il existe plusieurs types d'échelle de réponse[5], dont les plus utilisées sont : 1) l'échelle différentielle sémantique, 2) l'échelle de Guttman (appelée aussi analyse scalaire), 3) l'échelle visuelle analogue, et 4) l'échelle de type Likert. Selon le nombre d'ancrages ou d'intervalles, ces échelles sont de nature ordinale (moins de dix) ou continue (dix ou plus). À propos de l'échelle différentielle sémantique, il s'agit de mettre en opposition un concept tel que celui du locus interne ou du locus externe dans la prise de décision (–3 = locus externe à +3 = locus interne), dont le nombre d'ancrages ou d'intervalles le plus commun se situe entre 7 et 9 (DeVellis, 2011). Quant à l'échelle de Guttman, il est demandé au répondant (ou à l'intervenant) de dire jusqu'à quel degré il est en accord avec une succession d'attitudes ou de comportements proposés (DeVellis, 2011). Plus particulièrement, l'échelle de Guttman repose sur un principe hiérarchique où une réponse positive à une proposition donnée implique d'inclure la proposition de rang inférieur (p. ex., l'échelle de distance sociale de Bogardus[6]). Pour l'échelle visuelle analogue, le répondant sera invité à mettre une croix sur une ligne qui peut varier de 0 à 10 (p. ex., évaluation de la douleur où 0 = aucune douleur et 10 = la pire douleur). Pour ce qui est de l'échelle de Likert, l'évaluation porte en général sur le degré d'accord ou de désaccord du répondant à propos d'une attitude, d'un comportement ou d'un autre phénomène à l'étude, et les degrés sont généralement de cinq à sept (p. ex., 1 = pas du tout d'accord à 7 = tout à fait d'accord) (Netemeyer *et al.*, 2003). Clark et Watson (1995) énumèrent les ancrages les plus populaires : jamais à toujours, pas du tout à tout à fait, pas du tout comme moi à tout à fait comme moi.

Selon le contexte de l'étude, l'équipe pourra choisir un nombre pair d'intervalles pour l'échelle de Likert (p. ex., quatre ou six) plutôt qu'impair afin d'éviter les réponses neutres ou nommées « refuge » et ainsi forcer les réponses du répondant vers un pôle ou un autre de l'échelle de réponse (Netemeyer *et al.*, 2003). Le point milieu peut aussi représenter une forme d'évitement parce que le répondant ne comprend pas bien l'énoncé (Netemeyer *et al.*, 2003) ou encore il peut refléter un désintérêt à répondre (DeVellis, 2011). Pour l'évaluation d'attitudes ou d'opinions, notamment dans les sciences humaines et sociales, les échelles les plus communes sont celles de Likert ou visuelle analogue. Ces deux échelles se prêtent en général très bien à de nombreuses analyses statistiques, et ce, même si l'échelle de Likert est une échelle ordinale. DeVellis (2011) souligne que l'échelle de Likert correspond aux processus de base neurologiques chez

---

5. Pour une discussion sur les avantages et les inconvénients de l'utilisation de chacune d'elles, nous invitons le lecteur à consulter DeVellis, 2011, p. 85-99.
6. Pour de plus amples informations concernant cette échelle, consulter le site Web suivant : <http://www.definitions-de-psychologie.com/fr/definition/echelle-de-distance-sociale-de-bogardus.html>.

l'humain, où il est plus facile pour le participant de répondre à partir d'une séquence d'intervalles (représentés par un chiffre) et où chaque nombre, disposé sur une ligne, occupe un espace précis. L'ancrage de l'échelle ne reste donc pas une métaphore pour le répondant, mais bien une représentation visuelle précise, car les intervalles (ou chiffres) occupent un espace particulier sur une ligne donnée.

Le choix du nombre d'intervalles dans l'échelle de réponse est très important, comme susmentionné. Il doit se faire, comme pour la sélection des items, en tenant compte des caractéristiques de la population cible et du type d'évaluation. Les personnes ayant un problème d'attention ou un niveau de scolarité inférieur au secondaire peuvent rencontrer des difficultés à utiliser des échelles qui ont un trop grand choix de réponses ou pour lesquels il est difficile d'avoir des repères familiers. Clark et Watson (1995) stipulent que si l'on augmente le nombre d'intervalles, on nuit à la validité des réponses. Par exemple, une échelle en sept points ou intervalles pourrait être trop abstraite pour le répondant, car il y a trop d'intervalles et elle n'est pas commune avec ce que le répondant connaît (p. ex., un score de 1 à 10). Dans ce cas de figure, un nombre inférieur à sept intervalles comme cinq ou quatre pourrait être plus adéquat. Aussi, il est important que les personnes comprennent l'ensemble des ancrages sur une échelle de réponse (p. ex., la notion utilisée de *quelque peu* reste difficile à comprendre) ou encore qu'elles sachent bien distinguer les nuances entre les ancrages pour éviter les ambiguïtés possibles et des réponses erronées (DeVellis, 2011). Dans ce dernier cas, les chercheurs pourraient choisir d'utiliser des échelles avec des personnages sourire pour faciliter les réponses à une échelle de Likert en sept points dont les ancrages restent trop abstraits pour le répondant (p. ex., pas du tout d'accord à tout à fait d'accord). L'exemple de l'échelle de satisfaction des domaines de vie en est un très bon : une cote de 1 est représentée par le visage le plus souriant et une cote de 7 est représentée par le visage le plus triste (Mercier, 1994).

Toutes choses étant égales par ailleurs, il est préférable d'utiliser des échelles de mesure à plusieurs niveaux de réponse en vue d'obtenir une plus grande variation dans les réponses (plutôt qu'une échelle dichotomique) et ainsi détecter des différences intra-individuelles dans le temps (sensibilité au changement) et interindividuelles (entre des groupes). Une échelle qui présente de faibles variations dans ses réponses (c.-à-d. un nombre d'intervalles restreint) sera aussi une échelle qui se prêtera moins bien au calcul de corrélations avec des scores provenant d'autres outils de mesure (DeVellis, 2011).

1.2.3. Validation

Dans le champ de la psychométrie, la validation est un processus où les chercheurs déterminent le degré de confiance qu'ils accordent aux inférences associées aux réponses des personnes à un outil de mesure (Streiner et Norman, 2008). La validation est toujours inhérente à un groupe de personnes dans un contexte donné et elle reste un processus d'hypothèses à tester. Pour assurer la validation d'un outil de mesure, il est important d'évaluer ses qualités psychométriques (ou métrologiques) à travers les notions de validité et de fidélité. La notion de validité a pour objectif de préciser ce que l'outil de mesure évalue et avec quel degré d'exactitude il le fait (Anastasi, 1994, p. 131). Bernier et Pietrulewicz (1997) précisent que la validité d'un outil de mesure ne peut être étendue à l'ensemble des situations : la validité d'un outil de mesure est toujours reliée à une situation particulière. La notion de fidélité renvoie, quant à elle, à la reproductibilité des scores obtenus par les mêmes personnes lorsqu'un même outil de mesure leur est administré à des occasions différentes, ainsi qu'à la précision avec laquelle cet outil mesure certaines caractéristiques (Anastasi, 1994). On comprend alors à la lecture de la définition des notions de validité et de fidélité qu'elles peuvent présenter des éléments de chevauchement, notamment lorsqu'il s'agit de calculer la cohérence interne et la sensibilité au changement d'un outil de mesure.

Dans les prochains paragraphes, les divers types de validité seront présentés, dont les plus communs sont les validités apparente et de contenu, les validités de critère et de construit (DeVellis, 2011). D'ailleurs, cette division était déjà recommandée par Cronbach en 1955. Il faut comprendre ici que la notion de validité est divisée en plusieurs types, mais elle devrait être considérée comme unifiée plutôt que comme éclatée ou parcellisée. Par la suite, les types de fidélité, c'est-à-dire la cohérence interne, la fidélité test-retest, la fidélité interjuges ainsi que la sensibilité au changement, et l'établissement de valeurs de référence (ou normes) seront décrits. Enfin, les auteurs de ce chapitre proposeront une séquence à suivre pour évaluer les divers types de validité et de fidélité lors de la validation d'un outil de mesure.

1.3. Différents types de validité

1.3.1. Validité apparente (*face validity*) et validité de contenu (*content validity*)

Les validités apparente et de contenu sont les premières qualités psychométriques qu'un outil de mesure devrait posséder. Netemeyer *et al.* (2003) regroupent ces deux types de validité sous le vocable de validité de translation (*translation validity*), car toutes deux font référence à la manière dont

les items traduisent le concept à l'étude. Ces deux types de validité peuvent être évalués qualitativement, mais ils sont parfois vérifiés à l'aide d'analyses statistiques simples de nature descriptive. Ils reposent sur la représentativité des items d'un concept, et dans quelle mesure ces derniers permettent d'avoir une information adéquate et complète à propos du concept à l'étude. Il est à noter que cette évaluation qualitative (voire quantitative) n'est que très rarement mentionnée dans les articles scientifiques, car elle est jugée souvent comme « allant de soi ». En l'occurrence, il faut alors être prudent et tenter d'évaluer si les items et le choix des échelles de réponse semblent pertinents et représentent adéquatement le concept.

Les validités apparente et de contenu sont présentées souvent comme des notions interchangeables, alors que certains auteurs s'évertuent, à juste titre, à faire la distinction entre ces deux types de validité (Netemeyer *et al.*, 2003). En ce sens, DeVellis (2011) a quelques réserves concernant la validité apparente, car il y existe parfois un flou, à savoir pour qui les items représentent bien le concept évalué (répondants ou experts). C'est pourquoi les auteurs de ce chapitre ont tenté de bien distinguer ces deux types de validité, apparente et de contenu. La première différence se situe dans le fait que la validité apparente est évaluée par les répondants et non par les experts (Streiner et Norman, 2008) afin de réduire l'insatisfaction des répondants lorsqu'ils remplissent le questionnaire, mais aussi pour augmenter leur coopération et leur motivation à participer à l'étude. Netemeyer *et al.* (2003) précisent que la validité apparente est celle qui est évaluée directement par les répondants (p. ex., clarté des items et pertinence), alors que la validité de contenu fait appel à des experts du concept à l'étude. Pour ce qui est de la validité de contenu, le chercheur suit un ensemble d'étapes plus systématiques et rigoureuses qu'une évaluation qui porte juste *sur l'apparence* (DeVellis, 2011). Par exemple, les experts sélectionnés pourraient juger de la bonne représentation du concept par les items, si le chercheur leur a, au préalable, fourni une définition claire du concept à évaluer en en précisant les dimensions sous-jacentes. Autrement dit, la validité apparente se rapporte à des éléments pratiques, de clarté et d'appropriation de l'outil par les répondants, alors que la validité de contenu est un processus plus rigoureux et plus exhaustif qui renvoie à la cohérence du concept (et de ses dimensions) avec la théorie à laquelle il est associé, la formulation des items, les consignes de l'outil et le type d'échelle de réponse (Netemeyer *et al.*, 2003). Contrairement à la distinction que met en relief DeVellis (2011) et évoquée plus haut, la validité de contenu peut être à la fois évaluée par des experts et des non-experts. Netemeyer *et al.* (2003) suggèrent de considérer les étapes suivantes dans la validité de contenu : 1) bien définir le concept à évaluer ainsi que ses dimensions ; 2) utiliser la population générale et des experts pour générer des items ; 3) utiliser de nombreux juges pour la validité

de contenu en quantifiant les résultats sous forme d'échelle ; 4) examiner la représentation proportionnelle des items pour chacune des dimensions ; et 5) reporter les résultats de la validité de contenu lors de la publication de la validation de l'outil, permettant ainsi aux chercheurs d'évaluer la pertinence et la représentativité des items pour l'outil de mesure.

1.3.2. Validité de construit (*construct validity*)

La validité de construit porte sur la relation qui existe entre la définition théorique du concept qui est mesuré et l'outil de mesure qui a été conçu (DeVellis, 2011). Elle évalue donc si l'outil de mesure permet d'avoir une information adéquate sur le concept. En ce sens, la validité de contenu se distingue de la validité de construit, puisque la première renvoie à des inférences sur la construction de l'outil alors que la seconde renvoie à des inférences à partir des scores obtenus à l'outil (Tenopyr, 1977). La validité de construit va au-delà de la validité de contenu (appréciation subjective du contenu de l'outil), car elle permet de quantifier et de tester des relations qui devraient exister entre le concept mesuré et d'autres concepts. Il faut savoir qu'il n'existe pas de seuils objectifs pour démontrer que l'outil de mesure a une validité de construit solide (DeVellis, 2011). La validité de construit est bâtie en fonction des résultats des validités apparente, de contenu et de critère, mais aussi selon quatre autres types de validité : 1) par l'utilisation de groupes extrêmes ; 2) par convergence (aussi appelée la validité convergente) ; 3) par discrimination (aussi appelée la validité discriminante), et 4) par analyses factorielles. On comprend alors que la taille d'échantillon recommandée pour chaque type de validité dépendra avant tout du type d'analyse qui sera effectuée (p. ex., analyse factorielle, comparaison de moyenne). La validité de construit est toujours un processus continu, car de nouvelles hypothèses peuvent émerger. Elle sera souvent considérée comme la base fondamentale de tous les types de validité (Cronbach et Meehl, 1955 ; Streiner et Norman, 2008, p. 260). Dans les prochains paragraphes, nous aborderons les divers types de validité de construit.

Par l'utilisation de groupes extrêmes

Lorsqu'il utilise la méthode par groupes extrêmes, le chercheur vérifie que les résultats obtenus après l'administration de l'outil de mesure varient (cliniquement) selon l'appartenance à un groupe ou à un autre. Par exemple, pour démontrer qu'un outil de mesure permet d'appréhender le concept de dépression, l'outil sera administré à un groupe de personnes ayant reçu un diagnostic de dépression et à un autre groupe n'ayant pas ce diagnostic psychiatrique. L'hypothèse de recherche sera que les sujets

ayant un diagnostic de dépression obtiennent un score significativement différent (présence de symptômes dépressifs) de celui de ceux qui n'ont pas le diagnostic (absence de symptômes dépressifs). Lorsque cette méthode est choisie, il est important de concevoir adéquatement les groupes de comparaison afin de s'assurer qu'ils sont bien différents dès le départ. Ce dernier élément peut être la pierre d'achoppement de ce type de validité, car parfois, il est ardu de bien distinguer des groupes, voire de concevoir des groupes extrêmes (Streiner et Norman, 2008). Ce type de validité s'apparente à la validité prédictive (voir plus bas pour une définition). Toutefois, la validité par groupes extrêmes ne vise pas une prédiction dans le temps, mais plutôt une évaluation qui permet de distinguer deux groupes à un même moment (DeVellis, 2011). Les analyses statistiques seront d'ailleurs différentes pour ces deux types de validité. Pour la validité par groupes extrêmes, une analyse par comparaison de moyennes (p. ex., test *t* ou analyse de variance) ou une analyse discriminante (voir le chapitre 17 de cet ouvrage) pourra alors être utilisée, alors qu'une régression logistique (voir le chapitre 18 de cet ouvrage) pourra être envisagée dans le cas de la validité prédictive.

*Par convergence/Validité convergente (*convergent validity*)*

La mesure étalon (*gold standard*) est une mesure reconnue comme donnant une mesure exacte ou adéquate du concept à évaluer. Quand une mesure étalon n'est pas disponible, un autre outil évaluant le même concept pourra être utilisé pour évaluer la validité convergente. Autrement dit, la validité convergente permet de vérifier si les résultats obtenus à la suite de l'administration de l'outil de mesure à la population cible sont corrélés positivement à ceux obtenus avec d'autres outils qui mesurent des concepts semblables ou qui sont théoriquement en relation (Netemeyer *et al.*, 2003). Par exemple, les résultats d'un groupe de personnes à deux outils qui mesurent les symptômes de dépression, soit le Beck Depression Inventory (Beck, Steer et Brown, 1996) et l'échelle de dépression du Brief Symptom Inventory (Derogatis et Melisaratos, 1983), devraient être significativement corrélés. Il est à noter que si le coefficient de corrélation est trop élevé, alors cela signifie qu'on évalue des concepts quasi identiques. La question qui en découle alors est de savoir s'il est vraiment utile d'avoir deux outils de mesure qui évaluent le concept cible de façon très similaire (Streiner et Norman, 2008). En général, il est attendu que le coefficient de corrélation soit plus élevé que 0,30, mais inférieur à 0,90. Cette large étendue des coefficients de corrélation est relative au concept qui sera utilisé pour soutenir la validité convergente du concept (ou outil de mesure) à l'étude. Comme mentionné plus haut, il peut s'agir d'un concept très similaire, mais mesuré à partir d'un autre outil ou encore, d'un concept théoriquement en lien. Dans ce dernier cas,

une corrélation proche de 0,30 sera satisfaisante alors que pour un concept très similaire, une corrélation beaucoup plus élevée sera attendue, avec un coefficient de 0,50 au moins (mais plus bas que 0,90). Lorsque l'on utilise la méthode par convergence, le choix des concepts et outils qui permettent de les opérationnaliser est donc essentiel.

La validité convergente s'oppose à la validité divergente au sens où cette dernière permet d'évaluer une relation négative entre deux concepts (p. ex., l'estime de soi est corrélée négativement aux symptômes dépressifs). Parfois, certains outils de mesure ne sont pas disponibles dans la littérature scientifique pour évaluer des concepts similaires. Le calcul d'une matrice de corrélations multitraits multiméthodes pourrait être dans ce dernier cas une alternative intéressante (Streiner et Norman, 2008). Par exemple, en utilisant deux types de méthode d'échelle de réponse aux items (échelle de Likert et échelle visuelle analogue) et différentes dimensions d'un même concept telles que le climat, la liberté et la réalisation du concept plus large des valeurs de travail (Zaniboni *et al.*, 2010), il est possible de prendre connaissance de la relative force des coefficients de corrélation de cette matrice (Netemeyer *et al.*, 2003, p. 78). Dans la mesure où ces coefficients sont significatifs et positifs, notamment ceux mettant en lien une même dimension et deux méthodes différentes, une évidence de la validité convergente pourra alors être soutenue. Dans le cas où les dimensions conceptuelles ne proviennent pas du même concept (c.-à-d. aucun lien théorique), les résultats qui ne sont pas significatifs pour une même méthode peuvent alors appuyer la validité discriminante (voir plus bas pour une définition). On comprend que la matrice de corrélation multitrait multiméthode peut s'avérer intéressante et pratique pour observer les patrons de corrélation entre des dimensions conceptuelles (en lien ou pas sur le plan théorique) et en utilisant des méthodes différentes. Toutefois, la méthode multitrait multiméthode n'est pas toujours facile à réaliser, car des méthodes différentes pour évaluer un concept ne sont pas toujours disponibles (Streiner et Norman, 2008).

*Par discrimination/Validité discriminante (*discriminant validity*)*

Lorsqu'il utilise la méthode par discrimination ou la validité discriminante, le chercheur vérifie que les résultats obtenus à la suite de l'administration de l'outil de mesure ne sont pas corrélés à ceux obtenus à d'autres outils qui ne présentent pas de relation théorique ou qui mesurent des concepts très différents (p. ex., les intérêts professionnels n'ont pas de lien théorique étroit avec le sentiment d'efficacité dans la recherche d'emploi). Le coefficient de corrélation devrait donc se rapprocher d'une valeur de 0. La validité discriminante ne doit pas être confondue avec la validité divergente, car

cette dernière évalue deux concepts théoriquement opposés (on s'attend à obtenir une corrélation négative), alors que la validité discriminante évalue deux concepts qui n'ont pas de relation théorique (on s'attend à obtenir une corrélation nulle ou proche de zéro).

Par analyses factorielles

Lorsque l'analyse factorielle exploratoire est utilisée, le chercheur explore le nombre de dimensions conceptuelles de l'outil (voir le chapitre 22 de cet ouvrage). Le chercheur peut aussi vérifier si l'outil présente bien une ou plusieurs dimensions (définies *a priori*) et dans ce cas, il utilisera l'analyse factorielle confirmatoire. Streiner et Norman (2008) mentionnent que l'analyse factorielle s'avère utile et nécessaire quand il s'agit d'évaluer deux ou plusieurs dimensions ou facettes d'un même concept. D'ailleurs, Cronbach et Meehl (1955) mentionnent que l'analyse factorielle est une analyse utile et économique, car elle permet de diviser un concept en plusieurs dimensions qui prennent sens et ainsi d'entériner les hypothèses sur la multidimensionnalité du concept à l'étude. L'analyse factorielle est un type d'analyse statistique complexe qui permet de vérifier s'il existe une ou plusieurs dimensions au sein de l'outil et quels items de l'outil appartiennent à quelle dimension, c'est-à-dire avec quel facteur l'item est le plus corrélé. L'interprétation des résultats de ce type d'analyse est difficile à effectuer lorsqu'on a peu de notions statistiques. Il est toutefois important de comprendre que cette méthode permet d'explorer ou de confirmer si l'outil mesure une ou plusieurs dimensions conceptuelles. Les deux types d'analyses factorielles, exploratoire et confirmatoire, sont souvent indiquées pour renforcer la validité de construit d'un nouvel outil de mesure (Corbière *et al.*, 2009 ; Lecomte, Corbière et Laisné, 2006).

En bref, la validité de construit d'un outil de mesure est un long processus, qui requiert de la patience et un travail de longue haleine, car elle est estimée en fonction de méthodes diverses et d'une accumulation d'évidences provenant de sources variées.

1.3.3. Validité de critère (*criterion validity*)

La validité de critère est une autre forme de validité qui implique l'utilisation d'une mesure étalon (DeVellis, 2011). Le choix de la mesure étalon est crucial, car c'est elle qui sert de critère pour déterminer si l'outil a une bonne validité de critère ou non. Il existe deux types de validité de critère : la validité concomitante et la validité prédictive. La validité concomitante (*concurrent validity*) permet d'évaluer la relation ou la corrélation entre un outil de mesure et sa mesure étalon (ou critère). Le nouvel outil et la mesure

étalon sont administrés au même moment. La validité prédictive (*predictive validity*) cherche, quant à elle, à déterminer la valeur d'un outil de mesure pour prédire une mesure étalon évaluée ultérieurement[7] (Cronbach et Meehl, 1955; DeVellis, 2011). Elle permet d'évaluer si un outil de mesure permet de prédire le rendement à un critère. En l'occurrence, le critère est la variable qui fait l'objet de la prédiction. Par exemple, le critère à prédire pourrait être le fonctionnement de la personne à son poste de travail après s'y être absentée pour des raisons de santé mentale. Comme on peut s'en douter, il est ardu d'avoir accès à des mesures étalons en réadaptation au travail ou en psychologie du travail pour évaluer ce type de fonctionnement. Par conséquent, la validité de critère n'est pas fréquemment étudiée pour les outils de mesure, et ce, même si de nombreux chercheurs tentent de démontrer la validité prédictive d'outils de mesure qu'ils ont conçus. Dans le meilleur des cas, les coefficients statistiques (en général, des coefficients de corrélation pour la validité concomitante et des coefficients bêta d'analyses de régression pour la validité prédictive) qui permettent de soutenir la validité prédictive d'un outil de mesure restent modestes, voire faibles même si significatifs. L'alliance thérapeutique ou l'alliance de travail est un bon exemple de cette situation: elle permet d'expliquer seulement entre 8% et 12% de la variance des effets thérapeutiques, quelle que soit l'école de pensée en psychologie, mais elle demeure significative (Martin, Garske et Davis, 2000). Ceci étant dit, la validité de critère d'un outil de mesure, lorsque satisfaisante, a des retombées cliniques intéressantes parce qu'elle met en relation le résultat de l'outil de mesure et un critère d'intérêt pour l'équipe clinique. Pour résumer, ces deux types de validité de critère, concourante et prédictive, se distinguent par leur aspect temporel, mais ont comme force commune de mettre en relation le concept évalué par l'outil de mesure avec une mesure étalon reconnue dans le domaine à l'étude (DeVellis, 2011).

1.4. Différents types de fidélité

La fidélité d'un outil de mesure renvoie aux notions d'homogénéité et de stabilité dans le temps (DeVellis, 2011). Elle est d'ailleurs intimement liée à la population à laquelle on veut appliquer la mesure (Streiner et Norman, 2008); il existe autrement dit une forme d'interaction entre l'outil et la population ciblée. Sur le plan pratique, cela signifie que le score obtenu par l'outil de mesure ne devrait pas changer pour une population donnée (DeVellis, 2011). Il est à noter que la définition de ce concept de fidélité est

7. Il existe aussi la validité *post-dictive*, où la mesure étalon ou le critère est évalué avant la variable indépendante (Netemeyer, Bearden et Sharma, 2003, p. 76)

controversée, et ce, depuis les années 1930 ; d'ailleurs, les auteurs utilisent une terminologie parfois différente (p. ex., précision, accord, reproductibilité, répétition et cohérence). En fait, chaque notion fait référence à un type de fidélité particulier, ce qui sera d'ailleurs abordé dans les prochains paragraphes.

1.4.1. Cohérence interne (*internal consistency*)

Le calcul de la cohérence interne est une mesure de la fidélité, mais peut être aussi une mesure de validité. Anastasi (1994, p. 150) mentionne à ce propos : « Il est clair que les coefficients de cohérence interne, qu'ils proviennent d'items ou de sous-tests, sont essentiellement des mesures d'homogénéité. Parce qu'il contribue à caractériser l'univers de comportements ou le trait mesuré par le test, le degré d'homogénéité est relié d'une certaine mesure à la validité de construit. » Le calcul de la cohérence interne permet de faire ressortir à la fois la corrélation interitems et la corrélation de chaque item avec le score global de l'échelle (Streiner et Norman, 2008). Ce type de fidélité de l'outil de mesure permet d'apprécier l'homogénéité des items d'un concept ou d'un outil de mesure. Elle se base sur le principe que tous les items d'un outil sont des indicateurs d'un même concept et qu'il devrait donc exister une relation significative entre ces derniers. Trois items est le nombre minimal requis pour évaluer la cohérence interne d'une dimension conceptuelle ou d'un concept. Lorsqu'un outil comprend plusieurs dimensions, il sera attendu qu'une relation plus forte (ou plus élevée) sera obtenue entre les items d'une même dimension qu'avec les items d'une autre dimension (voir aussi la notion d'indice de saturation dans le chapitre 22 de cet ouvrage[8]. La cohérence interne est évaluée à l'aide du coefficient alpha de Cronbach (1951)[9] ou par le calcul de corrélations interitems. Selon Streiner et Norman (1995), le coefficient alpha de Cronbach devrait se situer entre 0,70 et 0,90 pour qu'il y ait une bonne homogénéité des items dans la dimension conceptuelle. DeVellis (2011) propose quelques repères pour le coefficient alpha de Cronbach :

$\alpha < 0{,}60$ = inacceptable ;

$0{,}60 \leq \alpha < 0{,}65$ = critique ;

$0{,}65 \leq \alpha < 0{,}70$ = minimalement acceptable ;

8. Clark et Watson (1995) font une différence entre la cohérence interne et l'homogénéité, où l'homogénéité ne fait pas référence à la multidimensionnalité d'un outil. Cependant, il nous semble justifiable d'avancer que l'homogénéité peut aussi être évaluée pour chacune des dimensions conceptuelles d'un outil de mesure.
9. Il est possible de calculer la cohérence interne pour des items de nature dichotomique grâce au calcul KR-20 (voir DeVellis, 2011), qui est équivalent au coefficient alpha de Cronbach.

$0,70 \leq \alpha < 0,80$ = acceptable ;

$0,80 \leq \alpha < 0,90$ = très bon ;

$0,90 \leq \alpha$ = l'échelle de mesure devrait être révisée, car on note une redondance de l'information.

Nous devons aussi savoir que le coefficient alpha de Cronbach est tributaire du nombre d'items appartenant à la dimension. Autrement dit, la valeur du coefficient alpha de Cronbach dépend non seulement de l'ampleur des corrélations entre les items, mais aussi du nombre d'items utilisés pour représenter le concept ou la dimension conceptuelle (Streiner et Norman, 2008). Si le nombre d'items est inférieur à cinq, il est fort possible d'obtenir un coefficient α autour de 0,65 (ce qui peut être acceptable dans ce cas), et lorsque le nombre d'items est élevé (c.-à-d. dix et plus), le coefficient α se situera vraisemblablement autour d'une valeur de 0,90. DeVellis (2011) donne un exemple intéressant à cet effet, où il présente la fluctuation du coefficient alpha de Cronbach selon le nombre d'items. Pour reprendre son exemple, si l'on considère quatre items intercorrélés à 0,40 en moyenne, la valeur de l'alpha sera égale à 0,80. Si on enlève un item (c.-à-d. si on a trois items au total), l'alpha sera d'une valeur de 0,75. Pour cinq items ayant le même coefficient de corrélation ($r = 0,40$), la valeur de l'alpha sera de 0,83 et si on a dix items, la valeur de l'alpha sera de 0,91. On comprend alors qu'il est important de considérer le nombre d'items dans l'interprétation du coefficient alpha de Cronbach. Streiner et Norman (2008) ajoutent que le coefficient alpha d'une valeur de 0,70 est considéré comme satisfaisant si le nombre d'items est inférieur à sept avec un échantillon de 100 sujets ; un coefficient alpha d'une valeur de 0,90 avec une échelle de plus de onze items sera aussi considéré comme satisfaisant.

Toutes choses étant égales par ailleurs, un coefficient alpha trop élevé ($\alpha > 0,90$) peut indiquer qu'il y a une redondance des items (Vallerand, 1989) et qu'il serait possible de retrancher certains items de la dimension conceptuelle. À noter qu'il est possible d'utiliser une fonction dans le logiciel statistique SPSS pour évaluer la valeur du coefficient alpha si on retranche un item de la solution initiale. Cette fonction peut être fort utile lorsque le chercheur souhaite réduire le nombre d'items de l'outil de mesure. Toutefois, dans certains contextes comme dans le milieu clinique, il est recommandé d'avoir un coefficient très élevé lorsque par exemple le professionnel de la santé souhaite réduire les erreurs pour poser un diagnostic (DeVellis, 2011 ; Streiner et Norman, 2008). Lorsque le coefficient est très faible ($\alpha < 0,60$), cela peut indiquer que les items ne mesurent pas le même concept, ou qu'ils mesurent des éléments très différents d'un même concept, ou encore que les sujets ou les évaluateurs cotent les items d'une même dimension de façon différente. Les auteurs de l'outil doivent alors évaluer si cette faible

corrélation est due à la formulation de l'item, aux critères d'observation ou à l'absence de relation entre l'item et le concept à évaluer. Dans ces cas de situation, l'équipe de recherche peut envisager de retrancher certains items de la dimension conceptuelle, de reformuler les items qui font défaut, voire d'en créer de nouveaux pour améliorer l'homogénéité de l'outil de mesure. Par contre, il faut être prudent lorsque le chercheur choisit d'obtenir un alpha plus satisfaisant en éliminant un item de l'outil de mesure. En effet, il n'est pas recommandé de laisser pour compte la validité de construit (p. ex., en éliminant un item de l'outil) dans le seul objectif d'augmenter la cohérence interne du concept (Clark et Watson, 1995). Par ailleurs, un des pièges d'une faible cohérence interne est que le chercheur en fasse fi et calcule un score global pour cette même dimension (addition des scores aux items). On comprend alors que ce score ne sera pas valide au premier sens du terme (on ne sait pas ce que l'on mesure), car il représente un score total d'éléments hétérogènes. Pour finir, l'analyse factorielle présente certaines similarités avec le calcul de la cohérence interne, car cette première tente de regrouper en des facteurs ou dimensions conceptuelles des items qui sont homogènes, c'est-à-dire présentant une corrélation interitems significative et positive[10].

1.4.2. Fidélité test-retest (*test-retest reliability*)

La fidélité test-retest évalue la stabilité dans le temps d'un outil de mesure. C'est pourquoi DeVellis (2011) nomme ce type de fidélité de *stabilité temporelle* (*temporal stability*). Lorsque l'on mesure le même construit plusieurs fois, on devrait obtenir le même résultat qu'à la fois précédente (ou un résultat similaire). La fidélité test-retest peut être évaluée en calculant un coefficient de corrélation (p. ex., coefficient de Pearson) afin de quantifier la variation qui existe entre les deux temps de mesure. Afin de pouvoir calculer ce coefficient de corrélation, plusieurs sujets (généralement au moins une vingtaine) devront être évalués deux fois. Il est important de s'assurer que l'état des personnes évaluées n'a pas changé entre les deux temps de mesure. Le délai entre les deux évaluations devrait être assez court : le test-retest peut se faire à quelques minutes d'intervalle pour un outil comme un thermomètre ou être de deux jours à deux semaines pour les intervalles les plus usuels dans les sciences humaines et sociales (Streiner et Norman, 2008). Ces intervalles de plusieurs jours permettent de contrôler les effets de mémoire chez le participant dans le cas d'une réponse à un questionnaire autorapporté. Le délai temporel entre le test et le retest peut d'ailleurs être très variable en

10. Le lecteur est invité à consulter Cortina, 1993, pour plus de détails sur la relation entre l'analyse factorielle et le coefficient alpha.

fonction de ce qui est mesuré et de la manière dont c'est mesuré. En effet, il faut se demander s'il peut y avoir un effet d'apprentissage, c'est-à-dire si la personne peut obtenir de meilleurs scores lors du retest (p. ex., une évaluation sur ordinateur où la personne pourrait être plus habile et rapide).

Selon DeVellis (2011), un faible niveau de fidélité peut être associé : *a*) à un réel changement du concept dans le temps (p. ex., phénomène qui évolue) ; *b*) au fait que le score du répondant à un questionnaire oscille dans le temps, selon le moment de la journée (p. ex., dû à la fatigue) ; *c*) à une erreur de mesure qui peut être inhérente au calibrage de l'outil ou à la compréhension des énoncés par le répondant ; *d*) aux caractéristiques des répondants (p. ex., besoin de l'individu de vouloir être cohérent dans le temps, apprentissage de la première passation). Il sera alors possible d'apporter des modifications à l'outil afin d'améliorer sa fidélité test-retest, à moins que l'outil ne soit considéré comme instable dans le temps et, dans ce dernier cas, la fidélité test-retest n'est pas appropriée (Vallerand, 1989). Lorsqu'un outil est administré par un évaluateur, la fidélité test-retest peut aussi porter le nom de fidélité intrajuge. La fidélité intrajuge est la stabilité dans le temps de la mesure lorsqu'elle est administrée par un même évaluateur à deux moments différents aux mêmes personnes.

Différents types d'analyses statistiques peuvent être utilisés pour évaluer ce type de fidélité test-retest. Les plus fréquemment rencontrés sont le calcul du coefficient de corrélation de Pearson[11]. Plus le coefficient de corrélation se rapproche de la valeur 1, plus la fidélité test-retest est jugée satisfaisante ; plus il se rapproche de la valeur 0, moins la fidélité est bonne. Un coefficient de corrélation se situant au-dessus de la valeur de 0,60 est généralement satisfaisant (Vallerand, 1989).

1.4.3. Fidélité interjuges (*inter-rater reliability*)

La fidélité interjuges évalue la stabilité d'un outil lorsqu'il est administré par plusieurs évaluateurs. Lorsque l'on mesure la même chose, au même moment, les évaluateurs devraient obtenir le même résultat s'ils utilisent le même outil de mesure. Si l'on n'obtient pas le même résultat, c'est qu'il y a une erreur de mesure qui peut être associée à une mauvaise définition des critères d'observation et de cotation de l'outil ou à un manque de formation des évaluateurs. La fidélité interjuges compare donc les résultats obtenus à la suite de l'administration du même outil de mesure à plusieurs personnes, au même moment, par deux ou plusieurs évaluateurs. La fidélité interjuges

11. De façon concrète, le coefficient de corrélation de Pearson est très proche du coefficient intraclasse (pour une illustration, voir Streiner et Norman, 2008, p. 183-184).

peut être évaluée en calculant le pourcentage d'accord entre les juges ou en calculant un coefficient de corrélation qui permet de quantifier la variation qui existe entre les juges. Afin de calculer ce pourcentage d'accord ou ce coefficient de corrélation, plusieurs sujets (généralement au moins une trentaine)[12] devront être évalués au même moment par deux juges ou plus. Pour ce faire, on utilise la technique du juge silencieux ou de l'enregistrement vidéo. Pour la technique du juge silencieux, un évaluateur administre l'outil de mesure et les autres évaluateurs donnent un score à l'outil de mesure sans intervenir. Comme il peut être intimidant ou impossible d'avoir plusieurs juges qui observent une personne au même moment, il est possible d'enregistrer l'évaluation sur vidéo ou bande audio. Les évaluateurs observent ou écoutent sur vidéo ou bande audio pour effectuer leur évaluation. La fidélité interjuges d'un outil de mesure peut être améliorée en essayant de comprendre les facteurs qui font que deux évaluateurs arrivent à des résultats différents. Une meilleure formation des évaluateurs ainsi qu'une meilleure définition des critères d'observation peuvent augmenter l'accord entre les juges (Streiner et Norman, 2008). La fidélité interjuges d'un outil peut s'avérer meilleure lorsque les juges viennent tout juste d'être formés à l'utilisation de l'outil. Par contre, elle peut se détériorer avec le temps à cause d'un manque d'application rigoureuse des critères d'évaluation.

Afin d'assurer le maintien de la fidélité interjuges, il peut être nécessaire de faire une mise à jour de la formation des évaluateurs ou de réévaluer régulièrement la fidélité interjuges de l'outil de mesure. L'accord interjuges est calculé grâce aux coefficients de corrélation intraclasse (ICC)[13] ainsi que par les coefficients Kappa (κ). Les coefficients de corrélation intraclasse ICC sont utilisés pour les niveaux de mesure par intervalles et de rapport. Les coefficients Kappa (équivalent au coefficient ICC [DeVellis, 2011, p. 50; Streiner et Norman, 2008]) sont utilisés pour les niveaux de mesure nominale/dichotomique et ordinale[14]. L'interprétation des coefficients Kappa et de corrélation intraclasse (ICC) est présentée dans le tableau 24.3 (Landis et Koch, 1977; Rosner, 2006).

12. Une table est proposée pour tenir compte de la marge d'erreur et de la probabilité d'accord (Gwet, 2010).
13. Le calcul de l'ICC a de nombreuses similarités avec celui de l'alpha de Cronbach (voir DeVellis, 2011, p. 44, pour une discussion). Aussi, Streiner et Norman (2008) suggèrent d'utiliser l'ICC plutôt que le coefficient Kappa, car il est plus souple. L'ICC permet d'isoler les facteurs qui affectent la fidélité et permet de contrôler pour les données manquantes. Il existe au moins six types de coefficients intraclasses (consulter Shrout et Fleiss, 1979).
14. Les coefficients Kappa peuvent aussi être utilisés pour les niveaux de mesure de type intervalle et auprès de plusieurs juges (pour une illustration, consulter Berry et Mielke, 1988).

Tableau 24.3.
Interprétation des coefficients Kappa et de corrélation intraclasse (ICC)

Coefficients Kappa	Coefficients intraclasses (ICC)
< 0 = désaccord	< 0,40 = fidélité médiocre
0 à 0,20 = accord très faible	0,40 à 0,75 = fidélité acceptable à bonne
0,21 à 0,40 = accord faible	> 0,75 = fidélité très bonne à excellente
0,41 à 0,60 = accord modéré	
0,61 à 0,80 = accord élevé	
0,81 à 1 = accord presque parfait	

1.5. Sensibilité au changement (*sensitivity to change*) et réactivité clinique (*responsiveness*)

La sensibilité au changement (*sensitivity to change*) est la qualité qu'a un outil de mesure pour détecter et quantifier un changement qu'il soit pertinent ou non, alors que la notion de réactivité clinique (*responsiveness*) permet d'évaluer si ce changement est cliniquement significatif (Streiner et Norman, 2008). Bien que plusieurs auteurs considèrent la sensibilité au changement et la réactivité clinique comme des attributs de la validité au sens large (Hays et Hadorn, 1992), Streiner et Norman (2008) les lient à la notion de fidélité. Pour évaluer la sensibilité au changement d'un outil de mesure, il est très important que sa validité et sa fidélité aient d'abord été appuyées. Si l'outil n'est ni valide ni fidèle, cela ne sert à rien d'évaluer sa sensibilité au changement ou sa réactivité clinique. Si l'outil est valide et fidèle, on comprendra alors que le changement qui est observé n'est pas dû, par exemple, à un problème de construit (validité de construit), à une faible cohérence interne (alpha de Cronbach) ou à manque de stabilité dans le temps (fidélité test-retest). Comme tous les outils de mesure ont une certaine erreur de mesure, le changement observé devra être plus grand que cette erreur de mesure. La sensibilité au changement et plus précisément la réactivité clinique d'un outil sont des conditions nécessaires pour évaluer les effets des interventions. Pour évaluer la réactivité clinique, il est important de cibler un critère qui permettra de savoir s'il y a eu véritablement un changement. Pour ce faire, on peut démontrer la sensibilité au changement ou la réactivité clinique d'un outil de mesure 1) en comparant les résultats de personnes dont l'état est stable aux résultats de personnes dont l'état a changé; 2) en comparant le changement obtenu avec l'outil de mesure au changement obtenu en utilisant une mesure étalon ou un critère (variable, autre outil); 3) en comparant le changement obtenu avec l'outil de mesure au changement observé par le clinicien (jugement clinique).

Ainsi, lorsqu'on évalue la sensibilité au changement et la réactivité clinique, l'outil de mesure doit être administré avant et après une intervention clinique, ce qui permet d'évaluer le changement de l'état de la personne. Différents types d'analyses statistiques peuvent être utilisés pour appuyer la sensibilité au changement. La plus simple est lorsque les auteurs vérifient s'il y a une différence significative entre les résultats obtenus au début du traitement et à la fin du traitement à l'aide d'un test *t* de Student pairé (test statistique qui évalue la différence entre deux moyennes) chez les sujets qui ont été classés comme ayant changé. Il est aussi possible de calculer la taille de l'effet (*effect size*). Lorsque le chercheur évalue la sensibilité au changement d'un outil de mesure, notamment sa réactivité clinique, il devrait rapporter les informations relatives au changement clinique minimalement attendu, c'est-à-dire de combien de points le score de l'outil de mesure devrait changer afin que l'on puisse affirmer que l'état de la personne s'est amélioré ou détérioré et n'est pas le fruit du hasard (Streiner et Norman, 2008). Une faible sensibilité au changement ne permettra pas de documenter les effets d'une intervention, ce qui la différencie d'ailleurs de la notion de réactivité clinique.

À la suite de l'étude minutieuse des qualités psychométriques de l'outil de mesure, notamment sa validité, sa fidélité et sa sensibilité au changement, il est alors possible d'établir les valeurs de références (ou normes) de l'outil de mesure. L'établissement de valeurs de référence ne représente pas une étape obligatoire dans la conception et la validation d'outils de mesure. Il peut s'avérer toutefois judicieux d'établir des valeurs de référence pour interpréter les résultats d'un individu à un test (p. ex., test cognitif, test d'intelligence) ou encore pour pointer un résultat critique à un outil de mesure en particulier tel que celui qui permet d'évaluer l'estime de soi de Nugent et Thomas (1993). En général, comme le souligne Vallerand (1989), le calcul des moyennes, des écarts-types, des percentiles, des scores z et T devraient suffire pour l'élaboration de valeurs de référence[15].

1.6. Séquence de l'évaluation des différents types de validité et de fidélité

Lorsque le concept à l'étude est opérationnalisé par des items et par une échelle de réponse appropriée, les auteurs du chapitre suggèrent de conduire en premier lieu les validités apparente et de contenu, pour ensuite procéder à l'évaluation de la validité de construit. La validité de construit par analyse factorielle nous semble justifiée à ce stade-ci pour faire ressortir ultérieurement les sous-dimensions conceptuelles, résultats qui peuvent d'ailleurs être soutenus par le calcul de la cohérence interne (alpha de Cronbach)

15. Pour des illustrations pratiques du calcul des percentiles, scores z et T, consulter Vallerand, 1989, p. 677.

de chacune des dimensions qui ont émergé de l'analyse factorielle exploratoire. Certains chercheurs suggèrent de calculer la cohérence interne des dimensions de l'outil avant même de faire l'analyse factorielle, ce qui nous semble peu utile, car si l'analyse factorielle nous renseigne sur de nouvelles dimensions conceptuelles, il faudra alors à nouveau faire le calcul de la cohérence interne des dimensions conceptuelles. L'analyse factorielle comme présentée dans le chapitre 22 de cet ouvrage est une méthode essentielle à maîtriser pour la validation d'un outil de mesure. Pour reprendre les mots de Vallerand (1989, p. 674) d'il y a un quart de siècle: «L'analyse factorielle fait maintenant partie de l'arsenal régulier du chercheur.» Il est donc important qu'un chercheur intéressé à la validation d'outils de mesure apprenne les tenants et aboutissants de l'analyse factorielle. En matière de séquences pour les analyses factorielles, Worthington et Whittaker (2006) recommandent dans tous les cas de situation de commencer par une analyse factorielle exploratoire pour ensuite faire une analyse factorielle confirmatoire (voir aussi le chapitre 22 de cet ouvrage). Commencer par une analyse factorielle exploratoire offre au chercheur une plus grande latitude pour peaufiner la solution factorielle de l'outil de mesure et, par conséquent, pour bien dégager et définir les dimensions sous-jacentes. L'analyse factorielle confirmatoire viendra de soi (avec un échantillon indépendant) pour confirmer ou infirmer la solution factorielle antérieurement observée.

Par la suite, lorsque ces facteurs ou ces dimensions conceptuelles sont bien consolidés, le chercheur pourra alors réaliser une analyse de validité de convergence (validité de construit) et de validité concomitante (validité de critère) pour s'assurer que les scores de l'outil de mesure sont corrélés avec des concepts similaires et une mesure étalon. En parallèle[16] peuvent être entreprises l'analyse de fidélité test-retest et celle de validité prédictive pour évaluer à la fois si ces dimensions sont stables dans le temps et si elles permettent de prédire un comportement mesuré ultérieurement. À la fin de cette séquence de validation, la sensibilité au changement et la réactivité clinique ainsi que l'établissement de valeurs de référence (ou normes) seront de mise pour bien assurer la validation d'un outil de mesure. Dans cette veine, la sensibilité au changement et les valeurs de référence sont intéressantes, car elles permettent d'effectuer des comparaisons avec d'autres groupes d'individus, notamment dans le cadre de projets de recherche d'envergure nationale ou internationale. Comme il peut être noté, tous les types de validité et de fidélité ne sont pas inclus dans cette séquence (p. ex., groupes extrêmes, validité discriminante), mais pourront aussi être considérés selon les intérêts du chercheur, les ressources dont il dispose, mais aussi les besoins du milieu clinique, pour ainsi compléter le tableau de l'étude de

16. Cette étape pourrait être réalisée ultérieurement si le chercheur n'a pas encore pu recueillir les données aux autres temps de mesure (p. ex. dans le cas d'une étude prospective ou longitudinale).

la validité et de la fidélité d'un outil de mesure. Pour finir sont consignées dans le tableau 24.2 les analyses les plus souvent utilisées pour évaluer les différents types de validité et de fidélité (p. ex., l'analyse de régression dans le cas de la validité prédictive). Pour ce qui est des échantillons requis pour rendre compte de ces types de validité et de fidélité, cela dépendra avant tout du type d'analyses effectuées. À titre de repère, il convient d'avoir un minimum de 20 sujets par groupe observé (comparaison de moyennes), de plus de 100 sujets pour les analyses factorielles (voir le chapitre 22 de cet ouvrage), d'un minimum de dix sujets pour chaque variable indépendante dans les analyses de régression (voir le chapitre 18 de cet ouvrage) et d'au moins 20 sujets pour les analyses de corrélation (p. ex., dans le cas de l'analyse de la fidélité test-retest et de la validité convergente).

2. TRADUCTION ET ADAPTATION TRANSCULTURELLE

La traduction et l'adaptation transculturelle d'un outil de mesure est un processus par lequel il est nécessaire de considérer à la fois les aspects de traduction (c.-à-d. linguistiques) et ceux reliés à l'adaptation contextuelle, notamment la culture (Beaton *et al.*, 2007). Dans le domaine de la recherche en psychologie du travail et de la réadaptation au travail notamment, les chercheurs se sont attachés à comprendre la présence ou l'absence d'un phénomène ou tout simplement à le décrire. Par exemple, ils peuvent évaluer les caractéristiques individuelles (santé physique et psychologique, personnalité, attitudes, comportements, etc.) ou les composantes d'une intervention (p. ex. l'alliance thérapeutique), leur influence sur le retour ou la réintégration au travail de personnes aux prises avec un trouble mental après une absence significative du marché du travail. Cette évaluation ne peut, par ailleurs, faire abstraction du contexte social, économique, culturel et politique dans lequel la personne évolue, elle est donc nichée dans cette réalité contextuelle. On comprend alors que l'évaluation d'un phénomène peut être réalisée à différents niveaux :

- *intraindividuel* : examiner les processus de stabilité ou de changement individuel dans le temps (p. ex., sentiment d'efficacité concernant la reprise de son travail), en tenant compte des événements de vie ;
- *interindividuel* : comparer des individus sur la base de certaines caractéristiques personnelles (p. ex., genre) ;
- *intergroupes* : comparer des personnes qui appartiennent à divers groupes et dont le sentiment d'appartenance peut aussi varier d'un individu à un autre selon ses expériences ;

- *transculturel et transcontextuel* : établir les similitudes et les différences interindividuelles et intergroupes de personnes qui vivent dans des contextes différents.

Dans les dernières décennies, avec l'ouverture au marché mondial, le dernier niveau d'analyse *transculturel et transcontextuel* est devenu de plus en plus essentiel sur le plan de la recherche. En effet, l'intensification des échanges économiques due à la mondialisation des marchés, les modes de communication soutenus par les technologies omniprésentes de fine pointe qui réduisent les distances géographiques, les phénomènes d'immigration qui accentuent la nécessité de prendre en compte les us et coutumes de plusieurs cultures, ne sont que quelques transformations de notre société qui mettent en exergue, par le fait même, le besoin d'une recherche de nature *transculturelle* ou *transcontextuelle*. Par ailleurs, avec l'hégémonie de la langue anglaise pour ce qui est de l'accessibilité aux articles de nature scientifique, il importe de les traduire dans la langue et le contexte qui sont appropriés. Dans le cas qui nous intéresse ici, il s'agit de traduire et d'adapter des outils en langue anglaise à la culture francophone dans laquelle l'outil va être utilisé (p. ex., québécoise, française, belge, suisse). Cette perspective comparative des individus provenant de diverses cultures ou contextes socioculturels, dont parfois les référents linguistiques sont différents, demande au chercheur de mettre en place une méthode rigoureuse, notamment pour la validité des données recueillies. Cette méthode consiste en premier lieu à une équivalence des outils de mesure d'une population à une autre, en considérant le contexte dans laquelle elle est nichée (langue, culture, valeurs, etc.).

2.1. Équivalence dans la recherche transculturelle

La recherche transculturelle est exposée non seulement à différents types d'erreurs ou de biais, mais aussi à des processus de mesure qui peuvent s'avérer artificiels (*artefact measurement*) (van de Vijver et Leung, 1997). Le premier type d'erreurs éventuelles est relié à la définition des notions conceptuelles à l'étude. Les concepts pourraient ne pas se superposer adéquatement d'une culture à une autre, par exemple, seulement certaines dimensions du concept peuvent être équivalentes d'une culture à une autre, uniques dans un contexte particulier ou encore nécessitent d'être scindées dans une culture. Dans ce dernier cas, par exemple, lors de la validation du questionnaire de Karasek (*job content questionnaire*) en langue vietnamienne, le concept intitulé « Latitude décisionnelle » exigeait d'être divisé en deux sous-dimensions conceptuelles pour assurer une bonne validité de construit par analyses factorielles (Hoang Thi *et al.*, 2013) : « Latitude décisionnelle – Autonomie »

et « Latitude décisionnelle – Autorité ». Ces auteurs font référence à l'histoire sociopolitique du Vietnam pour rendre compte de cette division conceptuelle dans les notions d'autorité et d'autonomie des travailleurs vietnamiens (Hoang Thi *et al.*, 2013). Un deuxième type d'erreurs possibles est relié à l'idiosyncrasie du contexte ou de la population à l'étude, comme le degré de familiarité avec un concept en particulier, la désirabilité sociale, la qualité de la relation entre la recherche et la communauté (voir le chapitre 26 de cet ouvrage), le niveau scolaire des individus pour ne citer que quelques exemples. Enfin, un troisième type d'erreurs concerne la formulation et la traduction des items (ou énoncés d'un questionnaire). Des items peuvent avoir une signification n'ayant aucune correspondance avec le contexte dans lequel la population est évaluée, comme la notion de schizophrénie en Afrique il y a de cela quelques décennies. En considérant ces sources possibles d'erreurs, une attention particulière doit être portée à la vérification de « l'équivalence » transculturelle ou transcontextuelle quand l'objectif du chercheur est de conduire une recherche dans un nouveau contexte ou pays.

2.2. Perspectives *emic* et *etic*

On note une influence significative du domaine de la linguistique sur celui de la recherche transculturelle, influence que l'on peut aborder par les perspectives dites *etic* et *emic*[17]. Berry (1989) a développé ces perspectives dans le domaine de la recherche transculturelle en psychologie. À titre de rappel, l'un des objectifs principaux de la psychologie est de mettre en relief des règles universelles de fonctionnement qui peuvent être généralisées (perspective *etic*), mais qui en même temps tiennent compte des spécificités contextuelles (perspective *emic*). La combinaison de ces deux perspectives doit être assurée en recherche transculturelle, et ce, en considérant les étapes méthodologiques et épistémologiques présentées ci-dessous.

1. Une approche *emic* permet de vérifier une hypothèse de recherche dans un contexte bien défini, par exemple, la mise en place d'accommodements de travail au Québec pour aider des personnes avec un trouble mental grave à se maintenir en emploi sur le marché ordinaire.
2. La même hypothèse, avec les mêmes outils de mesure (voir ci-dessous pour l'adaptation linguistique), peut être étudiée dans un autre contexte culturel ou un contexte organisationnel différent. On pense,

17. L'appellation *etic* et *emic* fait une référence sonore à la phonétique, le domaine de la linguistique qui étudie les sons communs aux différentes langues, et aux phonèmes, soit les sons propres à une langue.

par exemple, à la mise en place d'accommodements de travail au Québec ou en Italie pour aider des personnes avec un trouble mental grave à se maintenir dans une entreprise d'économie sociale. Cette double considération sur les plans culturel et linguistique renvoie à la notion d'*etic imposé*, pour reprendre la notion de Berry (1989).

3. Ce *passage* d'un contexte à un autre peut permettre de rendre compte des spécificités, des variations possibles dans l'interprétation et l'explication de l'hypothèse originale, des problèmes d'équivalence, et engendrer le besoin d'approfondir les résultats par l'utilisation d'une approche qualitative ou quantitative (approche *emic* dans le deuxième contexte, québécois ou italien).

4. La comparaison systématique entre les résultats obtenus dans les deux cas où la perspective *emic* est mise de l'avant (1 et 3) offre au chercheur une meilleure compréhension de l'effet du contexte (culturel, linguistique, économique, normatif, etc.) sur les résultats obtenus. Par exemple, il est possible d'observer que certains accommodements de travail sont propres au contexte québécois, tout en distinguant le milieu de travail (entreprise sociale ou marché ordinaire) ou propre au contexte culturel (italien ou québécois) dans un milieu organisationnel semblable (c.-à-d. entreprise sociale). Pour illustrer plus précisément cet exemple, la flexibilité des horaires de travail semble plus facile à mettre en place au Québec lorsque la personne occupe un emploi dans une entreprise d'économie sociale que sur le marché du travail ordinaire. Pour les différences entre l'Italie et le Québec, on note que des accommodements tels qu'une formation sur les problèmes de santé mentale est fournie plus souvent aux collègues de travail italiens que québécois. Cette différence peut s'expliquer en partie par le fait que les entreprises sociales italiennes incluent différents types d'employés présentant une vulnérabilité (p. ex., physique, cognitive, sensorielle), tandis que certains entreprises sociales au Québec sont particulièrement dédiées à une population ayant un trouble mental. Une hypothèse qui en découle est que les effectifs de certaines entreprises sociales québécoises sont déjà sensibilisés à cette thématique et, donc, une formation sur les problèmes de santé mentale offerte aux collègues de travail est moins pertinente en contexte québécois.

5. À partir des résultats de l'étape 4, on peut envisager une approche *etic dérivée* (*derived etic*) qui met en évidence les éléments communs des accommodements de travail entre les deux contextes, en l'occurrence le contexte culturel de l'entreprise d'économie sociale (Italie et Québec), et par conséquent mettre en lumière les règles de fonctionnement psychosocial qui restent stables quel que soit le contexte

culturel. Le même principe peut s'appliquer à deux contextes organisationnels (ordinaire ou protégé) dans une même province canadienne comme le Québec. À cette étape, il est donc possible d'étudier les similitudes et les différences quant à l'influence des accommodements de travail sur le maintien en emploi, dans au moins deux contextes. Par exemple, la mise en place de soutiens naturels par le supérieur immédiat et des collègues de travail est une mesure qui est transversale à tous les contextes, culturels et organisationnels (Corbière, Villotti *et al.*, sous presse-b).

2.3. Garantir l'équivalence à travers la traduction des outils

L'une des étapes méthodologiques nécessaires pour garantir l'équivalence des outils de mesure d'un contexte particulier à un autre est celle de la traduction et de l'adaptation culturelle de ces outils (Beaton *et al.*, 2007). Ainsi, l'équivalence entre les deux versions d'un outil, sur le plan sémantique, idiomatique et conceptuel, est conservée (Beaton *et al.*, 2007). Plusieurs auteurs suggèrent des étapes communes pour assurer une validation transculturelle d'un outil de mesure (Beaton *et al.*, 2007; Vallerand, 1989). Dans les prochains paragraphes, il s'agira de faire une synthèse de ces étapes, en les critiquant au besoin (voir aussi le tableau 24.4).

Tableau 24.4.
Étapes pour la traduction et l'adaptation transculturelle d'un outil de mesure

Étapes	Indicateurs des étapes de la traduction et de l'adaptation transculturelle : de la langue anglaise à la langue française
Traduction	▪ Deux traducteurs professionnels dont la langue maternelle est le français (l'un connaissant le domaine et l'autre pas) traduisent, de façon indépendante, le questionnaire (en anglais) dans la langue ciblée (le français).
	▪ La consigne pour remplir le questionnaire ainsi que les items et l'échelle de réponse sont traduits à cette étape-ci.
Synthèse	▪ Accompagnés par un médiateur (p. ex., un des chercheurs de l'équipe), les deux traducteurs font la synthèse des deux traductions en français et rapportent les écarts dans un rapport écrit.
	▪ Il est préférable de chercher le consensus entre les traducteurs plutôt qu'un compromis.
	▪ Une traduction française qui tient compte des résultats des deux traducteurs sera utilisée à l'étape suivante.

(*suite*)

Tableau 24.4. (*suite*)

Traduction inversée	▪ Deux traducteurs professionnels dont la langue maternelle est l'anglais traduisent de façon indépendante le questionnaire qui a été traduit en français dans la langue ciblée, soit l'anglais.
	▪ À cette étape, il est préférable que les deux traducteurs ne connaissent pas le domaine (ou le concept étudié) afin de mettre en relief les imperfections (ou biais) de traduction des traducteurs de l'étape 1.
	▪ Cette traduction inversée permet de relever les erreurs significatives dans les premières traductions.
	▪ Un rapport rédigé séparément par les deux traducteurs est fourni.
Révision par un comité expert	▪ La composition du comité expert comprend idéalement un chercheur qui connaît la méthode, les quatre traducteurs, un expert du domaine ciblé et si possible le concepteur du questionnaire qui a été traduit. ▪ Une révision de tous les rapports ainsi que des questionnaires traduits dans les deux langues (anglais et français) est effectuée par le comité expert. ▪ Chaque élément qui présente un écart ou qui fait consensus est revu. ▪ Tous les éléments qui ont amené une décision chez le comité devraient être consignés dans un rapport écrit. ▪ Une attention particulière sera apportée aux équivalences sémantique, conceptuelle, idiomatique et expérientielle. ▪ La production d'une version quasi finale du questionnaire en langue française est faite.
Prétest du questionnaire auprès de la population cible	▪ Une dizaine de personnes de la population cible remplissent le questionnaire (en français) en ayant comme consigne de mentionner s'ils comprennent bien chaque item, la consigne et l'échelle de réponse du questionnaire. Si toutefois des analyses quantitatives sont réalisées, alors un échantillon plus grand devrait être envisagé, soit 30 personnes et plus.
	▪ L'objectif est de relever les items présentant une ambiguïté, une incompréhension ou un autre problème éventuel.
	▪ Une révision du questionnaire se fait au besoin. ▪ La nouvelle version sera utilisée pour une évaluation de plus grande envergure (p. ex., projet de recherche).
Évaluation des qualités psycho-métriques de l'outil	▪ Types de validité et de fidélité présentés au tableau 24.2.
	▪ Analyses factorielles confirmatoires à deux ou plusieurs échantillons avec le calcul de l'invariance.

Note : Quand les mots *traduire* ou *traduction* sont utilisés, ils devraient être lus de la façon suivante : *traduire et adapter* ou *traduction et adaptation*. En effet, toute traduction implique une adaptation à un contexte particulier (population, culture, etc.).

La *première étape* vise la traduction initiale ; elle est conduite pour passer de la langue originale (c.-à-d. l'anglais) à la langue ciblée (c.-à-d. le français). Beaton *et al*. (2007) suggèrent qu'au moins deux traductions indépendantes (de deux traducteurs professionnels) soient faites pour ensuite les comparer en vue de dégager les écarts possibles[18]. En effet, comme Vallerand (1989) le souligne, il est fortement recommandé d'engager dans ce processus de traduction plus d'une personne, car les biais linguistiques, psychologiques et de compréhension d'une seule et même personne peuvent influencer significativement le produit final de la traduction. Ces écarts sont le reflet des ambiguïtés linguistiques sur lesquelles il faudra travailler. Les traductions sont réalisées par des traducteurs professionnels bilingues, mais dont la langue maternelle est la langue cible. Il est aussi recommandé qu'un premier traducteur ait une connaissance de la matière à traduire et que l'autre soit naïf concernant la matière. Le premier devrait garantir l'équivalence sur le plan du contenu et de la mesure ; le second traducteur devrait être plus attentif aux problèmes de compréhension du vocabulaire pour la population générale, tout en ayant à l'esprit les utilisateurs du questionnaire.

La *deuxième étape* consiste en la création d'une version de l'outil qui tient compte de la version originale et des deux traductions de la phase précédente. Cette opération devrait être réalisée par un troisième expert (p. ex., un des chercheurs de l'équipe), qui cherche des points de médiation en produisant un rapport sur le processus qui a conduit à la synthèse.

La *troisième étape* consiste en une traduction inversée (*back translation*) dans la langue originelle (dans notre cas l'anglais). Cette opération de retour à la langue originelle permet de vérifier la validité de la traduction en fait de fidélité du contenu des items. À cette étape-ci, il est aussi recommandé par Beaton *et al*. (2007) de consulter deux traducteurs professionnels. Cette étape est bien sûr la responsabilité de l'équipe de chercheurs qui entreprend une traduction et une adaptation transculturelle. La traduction inversée représente sans conteste des coûts élevés puisqu'il est requis de refaire la traduction par deux autres traducteurs (Caron, 2002 ; Haccoun, 1987). Si les deux traducteurs sont des professionnels (ce qui est recommandé, et non pas des collègues bilingues), le coût total de la traduction et de l'adaptation transculturelle ne va être que plus élevé. Par ailleurs, la traduction inversée, même si un rapport écrit est fourni, peut présenter quelques écueils importants, à savoir comment traiter des divergences linguistiques

18. Il importe de savoir que le travail de traduction pourra être différent selon qu'il s'agit de traduire de l'anglais au français dans un même pays comme le Canada ou encore de traduire de l'anglais d'Angleterre à la langue française de France. Dans le premier cas, on trouvera, toutes choses étant égales par ailleurs, plus de similarités sur le plan culturel. Bullinger *et al*. (1993) font référence à ces distinctions par les notions de nation et de culture.

jugées minimes; la question qui se pose alors est s'il faut recommencer tout le processus. Cette étape reste donc la décision du chercheur, prise en fonction notamment de la complexité des items à traduire et des ressources financières et humaines dont il dispose.

Une alternative intéressante à la traduction inversée et qui a été entreprise par notre équipe est, d'abord, que deux chercheurs bilingues de l'équipe traduisent le questionnaire pour ensuite réunir les deux traducteurs professionnels de l'étape 1 avec l'équipe de recherche afin de confronter les quatre traductions. La rencontre est certes plus longue pour échanger à propos des divergences (et convergences), mais elle permet de repérer les «anomalies» de la version originelle (phrase bancale, intervalle d'échelle boiteux, etc.). Cette alternative s'avère encore plus fructueuse lors de la conception de nouveaux outils de mesure par l'équipe de recherche (Corbière, Durand, St-Arnaud, Briand, Loisel et Fassier, 2011-2013).

La *quatrième étape* prévoit la constitution d'un comité expert pour l'analyse des détails des versions traduites de l'outil. Beaton *et al.* (2007) et Vallerand (1989) mentionnent plusieurs formes de comité, dont celui qui semble le plus solide est composé de l'équipe de chercheurs, des traducteurs impliqués dans le processus et du concepteur de l'outil. Dans le cadre d'une approche participative (voir le chapitre 26 de cet ouvrage), il sera question d'inclure une personne qui représente la population cible, mais aussi de déterminer à l'avance qui prendra la décision finale pour le choix d'un item plutôt qu'un autre (voir la notion *adjudicator* dans *ESS Round 5 Translation Guidelines*). Bullinger, Anderson, Cella et Aaronson (1993) précisent que les membres du comité devraient accorder autant d'importance à ce qui ressort des échanges qu'à la traduction comme telle. En ce sens, l'examen du comité sera centré sur les aspects décrits ci-dessous, dont certains ont déjà été abordés dans ce chapitre (Beaton *et al.*, 2007; Bullinger *et al.*, 1993; Flaherty *et al.*, 1988; van de Vijver et Leung, 1997).

- *L'équivalence sémantique*: Le sens des mots (p. ex., degré d'intensité), les ambiguïtés des mots utilisés, les difficultés grammaticales doivent être relevés. Le sens de chaque item est le même dans chaque culture après la traduction dans la nouvelle langue.
- *L'équivalence idiomatique*: La construction d'idiomes et d'un langage courant est jugée importante, mais il ne faut pas qu'elle fasse perdre le sens des items originels. La méthode de la traduction inversée est importante pour ce type d'équivalence, car il est parfois ardu de trouver des idiomes équivalents. Par ailleurs, certains de ces idiomes peuvent présenter une forte connotation émotive. Un francophone vivant au Québec, en Belgique, en Suisse ou en France utilise ses

propres expressions qui parfois sont identiques dans la forme, mais qui n'ont pas systématiquement le même sens (p. ex., le mot *sucette* («friandise») en France est l'équivalent de *suçon* au Québec, et le mot *sucette* («marque sur la peau») au Québec est l'équivalent de *suçon* en France).

- *L'équivalence fonctionnelle*: La version traduite de l'outil présente une signification semblable à la version originelle plutôt qu'une traduction littérale.
- *L'équivalence expérientielle*: La vérification que le contenu des items est applicable à l'expérience et à la vie quotidienne des sujets impliqués dans les contextes et la culture cible.
- *L'équivalence conceptuelle/de contenu* (qui fait référence aux validités de construit et de contenu): Le concept est équivalent dans chaque culture ou contexte. Le contenu de chaque item de l'outil est pertinent pour comprendre le phénomène dans chaque contexte ou culture à l'étude.
- *L'équivalence technique/opérationnelle*: La méthode d'évaluation (p. ex., entrevue, version papier) est comparable dans chaque culture concernant les données qui sont produites. Il est important de se soucier du format dans lequel l'outil de mesure a été conçu, car les modalités de présentation peuvent aussi avoir une influence sur la façon de répondre au questionnaire (p. ex., la présence ou l'absence d'un intervieweur). Ce type d'équivalence renvoie à l'équivalence métrique (p. ex., mêmes ancrage et intervalle). Sans ces derniers types d'équivalence de la mesure, plusieurs biais d'interprétation pourront être observés lors d'une comparaison internationale.
- *L'équivalence de critère* (qui fait référence à la validité de critère): L'interprétation de la mesure du concept reste la même lorsque comparée avec les normes de chaque culture à l'étude.

Tenir compte de l'ensemble de ces équivalences est essentiel lors de la traduction d'un outil de mesure, l'objectif ultime étant de préserver une signification la plus proche possible de la version originelle, mais de ne pas tomber dans le piège d'une traduction littérale sans prise de recul (Bullinger *et al.*, 1993). En ce sens, l'*équivalence de construit* doit être garantie quand on utilise le même outil de mesure dans deux ou plusieurs contextes ou pays. La première vérification renvoie à la capacité d'un outil à mesurer, sans équivoque, le même concept auprès de groupes de personnes vivant dans des contextes culturels différents. Cette vérification est essentielle, car dans la majeure partie des cas (pour ne pas dire tous), l'outil de mesure est

conçu tout d'abord dans un contexte particulier. Comme Gelfand, Leslie et Fehr (2008) le mentionnent sans retenue, l'hégémonie scientifique des pays comme les États-Unis et l'Angleterre dans certains domaines scientifiques tels que la psychologie organisationnelle n'est qu'un constat : « [*cross-cultural organizational psychology*] *remains a US export business* » (p. 494). Par exemple, le concept d'« engagement organisationnel » (*organizational commitment*) est d'origine américaine, et la plupart des études conduites sur ce sujet ont été réalisées dans des pays occidentaux. Toutefois, ce concept pourrait avoir diverses acceptions selon la culture dans laquelle il est évalué, notamment dans des pays où le travail est perçu comme un statut identitaire important ou encore comme une activité qui renvoie à la souffrance. Gelfand, Erez et Aycan (2007) ont mis en évidence que, même s'il y a une équivalence satisfaisante des outils de mesure en apparence, les facteurs prédictifs (*nomological network*) et les résultats (*outcomes*) de l'engagement organisationnel, pour poursuivre avec cet exemple, peuvent diverger d'un contexte à un autre.

En lien avec l'importance de la notion d'équivalence transculturelle et comme il a été abordé dans la première partie de ce chapitre, certains écueils lors de la traduction des items peuvent être évités tels que (non exhaustifs) : 1) le genre, notamment dans la langue française, qui ne doit pas être omis ; 2) l'omission de certains mots ou autres subtilités telles les notions d'espace et de temps ; 3) les conjonctions de coordination (p. ex., *et*, *ou*, *car*) doivent aussi être considérées attentivement pour ne pas modifier l'articulation des idées ; 4) la séquence des idées doit être respectée lors de la traduction ; 5) la terminologie utilisée (p. ex., éléments techniques) doit être comprise par les personnes cibles qui répondront au questionnaire ; 6) l'importance de faire preuve de parcimonie en n'ajoutant pas de mots inutiles lors de la traduction ; 7) les étiquettes des intervalles de l'échelle doivent être similaires d'une langue à une autre ; 8) la présentation du format du questionnaire est semblable[19].

La *cinquième étape* consiste en un prétest de la nouvelle version de l'outil sur un petit groupe d'individus (d'une dizaine à plusieurs dizaines selon les analyses). Les personnes impliquées dans la validation devraient répondre au questionnaire et, ensuite, donner des indications qualitatives sur leurs perceptions concernant le contenu des items, leur sens et leurs ambiguïtés. Cette étape est cruciale, car elle permet de vérifier les interprétations erronées des items ou d'autre matériel du questionnaire (c.-à-d. consigne, échelle de réponse) et ainsi assurer une validité de construit

19. Nous invitons le lecteur à prendre connaissance des recommandations de *ESS Round 6 Translation Guidelines :* <http://www.europeansocialsurvey.org/docs/round6/methods/ESS6_translation_guidelines.pdf>.

solide (Beaton *et al.*, 2007). Vallerand (1989) propose deux techniques à ce propos, celle de Schuman (1966) et celle de Mitchell (1966). Dans la première (Schuman, 1966), il s'agit de prendre de façon aléatoire certains items du questionnaire et de demander à des personnes de nous renseigner sur leur compréhension. Dans la seconde technique (Mitchell, 1966), un groupe d'individus au nombre de 20 sont interrogés à un moment donné sur leur compréhension des énoncés et à une période ultérieure, ils sont à nouveau questionnés. Dans les deux cas, s'il y a une incompréhension ou un décalage dans les réponses, c'est que les items présentent des ambiguïtés et qu'ils devraient donc être retravaillés. Une autre façon serait de demander aux répondants d'encercler les items qui leur semblent problématiques. Haccoun (1987) suggère et compare trois techniques qui permettent de vérifier l'équivalence de la traduction des items en suivant une approche statistique : la technique de Prince et Mombour (1967), la technique par analyses des items (Hulin, Drasgow et Komocar, 1982) et la technique par groupe unique (Haccoun, 1982). Cette dernière technique semble la plus simple et la plus rigoureuse des trois exposées. Il est demandé à un groupe d'individus bilingues de répondre à deux questionnaires (anglais et français) et à partir des réponses obtenues, des corrélations sont établies entre les deux versions (Haccoun, 1982). Toutefois, comme l'auteur le mentionne, une limite vient du degré de bilinguisme des participants à l'exercice, mais il existe aussi deux autres limites à cette technique, soit de l'ordre de l'effet de mémoire des participants et de la stabilité du concept dans le temps, qui est difficile à évaluer dans ce cas de figure. Pour contrecarrer la première limite, Vallerand (1989) propose une évaluation du bilinguisme sous forme de pointage. Pour ce qui est des deux autres limites, il est difficile de les contourner, car comme il est mentionné dans la première partie de ce chapitre, la fidélité test-retest n'est jamais parfaitement atteinte et un coefficient de cohérence interne de 0,70 est jugé satisfaisant. D'où la difficulté de se fier aux résultats de la méthode de Haccoun (1982), car ils ne resteront qu'approximatifs. Une fois analysées, les indications du prétest devraient révéler qu'aucun problème majeur n'a été relevé dans la version traduite. Si ces derniers critères sont satisfaisants, on peut alors envisager d'utiliser l'outil de mesure dans des recherches de plus grande envergure.

La *sixième étape* consiste en l'évaluation des qualités psychométriques de l'outil de mesure, ce qui correspond à l'étape de validation d'un outil de mesure, abordée dans la première partie de ce chapitre. Toutefois, il existe un exercice intéressant supplémentaire lors de la traduction et de l'adaptation transculturelle d'un outil de mesure, lequel vise à tester l'équivalence des matrices de covariance entre deux échantillons provenant de contextes différents, et ce, grâce à l'utilisation d'analyses factorielles confirmatoires par calcul d'invariance. Le chercheur utilisant l'approche confirmatoire peut

donc évaluer statistiquement si les solutions factorielles issues des versions originales et traduites sont équivalentes (Vallerand, 1989). Dans la mesure où les deux versions ne sont pas équivalentes (p. ex., les coefficients de corrélation des dimensions conceptuelles, les indices de saturation des énoncés à une dimension conceptuelle)[20], ces informations peuvent être reprises par le comité en vue de reformuler les items insatisfaisants. En outre, si l'on trouve des ambiguïtés dans la traduction d'un outil, la tendance est d'éliminer les items problématiques. En procédant ainsi, on risque de nuire à de possibles différences ou nuances culturelles ou de ne pas tenir compte de la spécificité d'un contexte par rapport à un autre (approche *emic* déjà décrite). S'arrêter à une parfaite équivalence des outils de mesure sur le plan statistique signifie que le chercheur adopte une approche « pseudo-*etic* » qui réduit exclusivement l'analyse transculturelle à la partie commune des deux cultures aux contextes à l'étude.

Cette seconde partie de chapitre a permis d'appréhender les éléments essentiels de la traduction et de l'adaptation transculturelle d'un outil de mesure. Le lecteur est invité à consulter les références de ce chapitre pour compléter ses connaissances. Pour finir, nous ne mettrons jamais assez en garde les chercheurs qui entreprennent une traduction maison d'un outil de mesure sans suivre une méthode rigoureuse, ce qui pourrait causer éventuellement des erreurs lors de l'analyse des données recueillies. Mais comment savoir qu'un outil de mesure a été traduit et adapté à une nouvelle culture ? Pour répondre à cette question, Corbière (2011) propose de créer un forum sur le Web pour échanger à propos des outils de mesure dans un domaine particulier (p. ex., psychologie du travail, réadaptation au travail) et d'élaborer des grilles informatisées pour chaque outil, mises à jour régulièrement (p. ex., traduction en cours, qualités psychométriques). Le travail de réseautage paraît donc essentiel pour prendre connaissance des travaux de traduction et d'adaptation d'un outil de mesure dans un contexte particulier.

CONCLUSION

Le processus de conception et de validation consiste en plusieurs étapes et exige temps et patience chez ceux qui décident de l'entreprendre. À l'issue de ce processus rigoureux, le chercheur devrait être en mesure d'apprécier les qualités psychométriques de l'outil de mesure, pour ainsi bien connaître ses forces et ses faiblesses tout en utilisant les résultats de façon adéquate. Il va de soi que tous les outils devraient avoir une bonne validité ainsi qu'une

20. À titre d'illustration, le lecteur est invité à consulter l'article de Corbière *et al.* (2006) sur ce type d'équivalences statistiques par l'utilisation d'analyses factorielles confirmatoires avec calcul d'invariance.

bonne fidélité avant leur utilisation. Parfois, à la lecture d'articles scientifiques, les étapes de conception et de validation ne sont pas toujours décrites et il existe aussi, comme nous avons pu le constater, différentes définitions de ces concepts dans la littérature scientifique, notamment les définitions de validité de construit et de validité de critère. Toutefois, les définitions retenues dans ce chapitre sont celles pour lesquelles il semble y avoir un plus grand consensus chez les experts dans le domaine.

Il est important de se rappeler qu'il n'existe pas d'outils de mesure parfaits, mais qu'ils sont tous perfectibles et que le processus de conception et de validation d'un outil de mesure est un travail ardu qui prend beaucoup de temps et d'énergie. Alors, avant de décider de concevoir un nouvel outil de mesure, il est recommandé de prendre connaissance de la littérature scientifique afin de s'assurer qu'un outil similaire n'existe pas déjà. Concevoir un nouvel outil de mesure est sans conteste un travail de longue haleine et une aventure en soi, car de nouvelles hypothèses et inférences jalonnent le processus exigeant à mettre en place pour les vérifier, notamment lorsque l'outil est modifié ou évalué dans d'autres contextes ou auprès de nouvelles populations. Ce travail est du même ordre pour ce qui concerne la traduction et l'adaptation transculturelle, car il est tentant de traduire un outil qui semble bien validé dans un pays sans être précautionneux des adaptations nécessaires dans un autre contexte donné. Prudence, rigueur, flexibilité, créativité et persévérance sont probablement des qualités que devrait posséder un chercheur qui s'intéresse à la conception et à la validation d'un outil de mesure. Quant à la question récurrente des étudiants de savoir à quel moment s'arrête la validation transculturelle d'un outil de mesure, nous vous laissons le loisir d'y trouver une réponse claire…

RÉFÉRENCES

ANASTASI, A. (1994). *Introduction à la psychométrie*, Montréal, Guérin Éditeur.

BEATON, D., C. BOMBARDIER, F. GUILLEMIN et M.B. FERRAZ (2007). *Recommendations for the Cross-Cultural Adaptation of the DASH & QuickDASH Outcome Measures*, Toronto, Institute for Work & Health.

BECK, A.T., R.A. STEER et G.K. BROWN (1996). *Beck Depression Inventory: Manual*, 2e éd., San Antonio, The Psychological Corporation.

BENSON, J. et F. CLARK (1982). « A guide for instrument development and validation », *American Journal of Occupational Therapy*, vol. 36, no 12, p. 789-800.

BERNIER, J.-J. et B. PIETRULEWICZ (1997). *La psychométrie: traité de mesure appliquée*, Montréal, Gaëtan Morin.

BERRY, J.W. (1989). « Imposed etics-emics-derived etics : The operationalization of a compelling idea », *International Journal of Psychology*, vol. 24, nº 6, p. 721-735, <doi : 10.1080/00207598908247841>.

BERRY, K.J. et P.W. MIELKE (1988). « A generalization of Cohen's Kappa Agreement Measure to interval measurement and multiple raters », *Educational and Psychological Measurement*, vol. 48, p. 921-933, <doi : 10.1177/0013164488484007>.

BULLINGER, M., R. ANDERSON, D. CELLA et N. AARONSON (1993). « Developing and evaluating cross-cultural instruments from minimum requirements to optimal models », *Quality of Life Research*, vol. 2, nº 6, p. 451-459.

CARON, J. (2002). *Un guide de validation transculturelle des instruments de mesure en santé mentale*, <http://instrumentspsychometriques.mcgill.ca/instruments/guide.htm>, consulté le 6 janvier 2012.

CLARK, L.A. et D. WATSON (1995). « Constructing validity : Basic issues in objective scale development », *Psychological Assessment*, vol. 7, nº 3, p. 309-319.

CONTANDRIOPOULOS, A.-P., F. CHAMPAGNE, L. POTVIN, J.-L. DENIS et P. BOYLE (2005). *Savoir préparer une recherche : la définir, la structurer, la financer*, Montréal, Gaëtan Morin.

CORBIÈRE, M. (2011). « L'utilisation d'outils de mesure en milieu clinique : mythe ou nécessité ? », *Le partenaire*, vol. 20, nº 3, p. 4-7.

CORBIÈRE, M., E. BROUWERS, N. LANCTÔT et J. VAN WEEGHEL (sous presse-a). « Employment specialists competencies for supported employment programs », *Journal of Occupational Rehabilitation*.

CORBIÈRE, M., M.-J. DURAND, L. ST-ARNAUD, C. BRIAND, P. LOISEL et J.B. FASSIER (2011-2013). *Validation du questionnaire « Obstacles au retour au travail et sentiment d'efficacité pour les surmonter » (ORTESES) auprès de travailleurs avec un trouble musculosquelettique ou un trouble mental*, Montréal, Institut de recherche Robert-Sauvé en santé et sécurité du travail (IRSST).

CORBIÈRE, M., F. FRACCAROLI, V. MBEKOU et J. PERRON (2006). « Academic self-concept and academic interest measurement : A multi-sample European study », *European Journal of Psychology of Education*, vol. 21, nº 1, p. 3-15, <doi : 10.1007/bf03173566>.

CORBIÈRE, M., N. LANCTÔT, N. SANQUIRGO et T. LECOMTE (2009). « Evaluation of self-esteem as a worker for people with severe mental disorders », *Journal of Vocational Rehabilitation*, vol. 30, p. 87-98.

CORBIÈRE, M., P. VILLOTTI, T. LECOMTE, G. BOND, A. LESAGE et E. GOLDNER (sous presse-b). « Work accommodations and natural supports for maintaining employment », *Psychiatric Rehabilitation Journal*.

CORTINA, J.M. (1993). « What is coefficient alpha ? An examination of theory and applications », *Journal of Applied Psychology*, vol. 78, nº 1, p. 98-104, <doi : 10.1037/0021-9010.78.1.98>.

CRONBACH, L.J. (1951). « Coefficient alpha and the internal structure of tests », *Psychometrika*, vol. 16, nº 3, p. 297-334.

CRONBACH, L.J. et P.E. MEEHL (1955). « Construct validity in psychological tests », *Psychological Bulletin*, vol. 52, p. 281-302.

DEROGATIS, L.R. et N. MELISARATOS (1983). « The Brief Symptom Inventory : An introductory report », *Psychological Medicine*, vol. 13, nº 3, p. 595-605.

DEVELLIS, R.F. (2011). *Scale Development : Theory and Applications*, Thousand Oaks, Sage Publications.

FLAHERTY, J.A., F.M. GAVIRIA, D. PATHAK, T. MITCHELL, R. WINTROB, J.A. RICHMAN et S. BIRZ (1988). « Developing instruments for cross-cultural psychiatric research », *Journal of Nervous and Mental Disease*, vol. 176, nº 5, p. 257-263.

GELFAND, M.J., M. EREZ et Z. AYCAN (2007). « Cross-cultural organizational behavior », *Annual Review of Psychology*, vol. 58, p. 479-514, <doi : 10.1146/annurev.psych.58.110405.085559>.

GELFAND, M.J., L.M. LESLIE et R. FEHR (2008). « To prosper, organizational psychology should… adopt a global perspective », *Journal of Organizational Behavior*, vol. 29, nº 4, p. 493-517, <doi : 10.1002/job.530>.

GWET, K.L. (2010). *Handbook of Inter-Rater Reliability*, 2e éd., Gaithersburg, Advanced Analytic Press.

HACCOUN, R.R. (1982). « La recherche explicative sur le terrain », dans Y. Bordeleau, L. Brunet, R.R. Haccoun, A.J. Rigny et A. Savoie (dir.), *Comprendre l'organisation : approches de recherches*, Montréal, Les Éditions Agence d'Arc.

HACCOUN, R.R. (1987). « Une nouvelle technique de vérification de l'équivalence de mesures psychologiques traduites », *Revue québécoise de psychologie*, vol. 8, nº 3, p. 30-39.

HAYS, R.D. et D. HADORN (1992). « Responsiveness to change : An aspect of validity, not a separate dimension », *Quality of Life Research*, vol. 1, nº 1, p. 73-75.

HOANG THI, G., M. CORBIÈRE, A. NEGRINI, M.K. PHAM et D. REINHARZ (sous presse). « The validation of the JCQ-K to measure occupational stress in Vietnam », *Psychological Reports*.

HULIN, C.L., F. DRASGOW et J. KOMOCAR (1982). « Applications of item response theory to analysis of attitude scale translations », *Journal of Applied Psychology*, vol. 67, nº 6, p. 818-825, <doi : 10.1037/0021-9010.67.6.818>.

LANDIS, J.R. et G.G. KOCH (1977). « The measurement of observer agreement for categorical data », *Biometrics*, vol. 33, nº 1, p. 159-174.

LECOMTE, T., M. CORBIÈRE et F. LAISNÉ (2006). « Investigating self-esteem in individuals with schizophrenia : Relevance of the Self-Esteem Rating Scale-Short Form », *Psychiatry Research*, vol. 143, nº 1, p. 99-108, <doi : 10.1016/j.psychres.2005.08.019>.

MARTIN, D.J., J.P. GARSKE et M.K. DAVIS (2000). « Relation of the therapeutic alliance with outcome and other variables : A meta-analytic review », *Journal of Consulting and Clinical Psychology*, vol. 68, nº 3, p. 438-450.

MERCIER, C. (1994). « Improving the quality of life of people with severe mental disorders », *Social Indicators Research*, vol. 33, nº 1, p. 165-192, <doi : 10.2307/27522779>.

MITCHELL, R. (1966). « The problems and possibilities of measuring social attitudes in African social surveys », communication présentée à la 9e conférence annuelle de l'Association pour les études africaines, Bloomington, Indiana.

Netemeyer, R.G., W.O. Bearden et S. Sharma (2003). *Scaling Procedures. Issues and Applications*, Thousand Oaks, Sage Publications.

Nugent, W. et J. Thomas (1993). « Validation of a clinical measure of self-esteem », *Research and Social Work Practice*, vol. 3, p. 208-218.

Prince, R. et W. Mombour (1967). « A technique for improving linguistic equivalence in cross-cultural surveys », *International Journal of Social Psychiatry*, vol. 13, nº 3, p. 229-237, <doi: 10.1177/002076406701300308>.

Rosner, B. (2006). *Fundamentals of Biostatistics*, 6e éd., Belmont, Duxbury Press.

Schuman, H. (1966). « The random probe : A technique for evaluating the quality of closed questions », *American Sociological Review*, vol. 31, p. 218-222.

Shrout, P.E. et J.L. Fleiss (1979). « Intraclass correlations : Using in assessing raters reliability », *Psychological Bulletin*, vol. 86, nº 2, p. 420-428.

Streiner, D.L. et G.R. Norman (1995). *Health Measurement Scales : A Practical Guide to Their Development and Use*, 3e éd., Oxford, Oxford University Press.

Streiner, D.L. et G.R. Norman (2008). *Health Measurement Scales : A Practical Guide to Their Development and Use*, 4e éd., Oxford, Oxford University Press.

Tenopyr, M.L. (1977). « Content-construct confusion », *Personnel Psychology*, vol. 30, nº 1, p. 47-54, <doi: 10.1111/j.1744-6570.1977.tb02320.x>.

Vallerand, R.J. (1989). « Vers une méthodologie de validation transculturelle de questionnaires psychologiques : implications pour la recherche en langue française », *Psychologie canadienne*, vol. 30, nº 4, p. 662-689.

van de Vijver, F.J.R. et K. Leung (1997). *Methods and Data Analysis for Cross-Cultural Research*, Thousand Oaks, Sage Publications.

Worthington, R.L. et T.A. Whittaker (2006). « Scale development research : A content analysis and recommendations for best practices », *The Counseling Psychologist*, vol. 34, nº 6, p. 806-838, <doi: 10.1177/0011000006288127>.

Zaniboni, S., M. Corbière, F. Fraccaroli et J. Perron (2010). « Work values of people with severe mental disorders registered in vocational programs : Validation of the Work Values Questionnaire », *Canadian Journal of Community Mental Health*, vol. 29, nº 1, p. 107-122.

CHAPITRE 25

LES MÉTHODES DE RECHERCHE MIXTES
Illustration d'une analyse des effets cliniques et fonctionnels d'un hôpital de jour psychiatrique

Catherine Briand
Nadine Larivière[1]

FORCES

- Elles permettent de répondre à des questions de recherche nécessitant plusieurs sources de données.
- Elles permettent de mettre au service d'une question de recherche les méthodologies de recherche des orientations quantitatives et qualitatives.
- Elles permettent au sein d'un même projet de recherche d'obtenir une richesse d'informations et d'approfondir le sujet de l'étude.

LIMITES

- Elles impliquent un processus de recherche qui peut être long (surtout les devis séquentiels) et qui nécessite des ressources importantes pour recueillir et analyser les données.
- Elles impliquent d'avoir des équipes de recherche avec multiexpertises : experts en méthodes quantitatives, qualitatives et mixtes.
- Elles peuvent amener des défis de publication et de synthèse de l'ensemble des résultats dans un même article scientifique.

1. Les auteures tiennent à remercier M[me] Véronique Thibault pour sa contribution à ce chapitre.

Les méthodes de recherche mixtes sont de plus en plus documentées, connues et utilisées. Elles s'inscrivent dans un troisième courant ou paradigme de recherche dans lequel on incorpore au sein d'une même étude (dans sa conceptualisation, son devis ou son type de données recueillies et analysées) des composantes des méthodes quantitatives et qualitatives (Creswell et Plano Clark, 2011 ; Johnson, Onwuegbuzie et Turner, 2007 ; Michaud et Bourgault, 2010). Les méthodes de recherche mixtes découlent de l'idée de la triangulation des résultats de recherche, c'est-à-dire que plusieurs résultats d'expérimentations différentes sont combinés pour mieux cerner un phénomène et assurer la validité des conclusions. Déjà dans les années 1950-1960, certains auteurs parlaient de « *multiple operationalism* » ou de « *between-methods triangulation* » pour nommer cette combinaison de méthodes et d'expérimentations, et pour penser autrement la production des nouvelles connaissances (Johnson *et al.*, 2007).

Dans ce chapitre, les neuf principaux devis issus des méthodes de recherche mixtes seront définis. Plus précisément, dans la première partie de ce chapitre, les raisons qui expliquent le choix des devis mixtes, les forces et les défis de chacun, et des suggestions de stratégies pour assurer une recherche de qualité seront abordés. Dans la seconde partie de ce chapitre, le devis mixte séquentiel explicatif fera l'objet d'une illustration. Le projet de recherche présenté est une analyse des effets cliniques et fonctionnels d'un hôpital de jour psychiatrique.

1. DÉFINITION ET CLASSIFICATION DES MÉTHODES DE RECHERCHE MIXTES

Johnson et collègues définissent les méthodes de recherche mixtes comme « un type de recherche dans lequel un chercheur ou une équipe de recherche combine des aspects des méthodes qualitatives et quantitatives (c.-à-d. les postulats, les outils de collecte de données, l'analyse, les techniques d'inférence) à des fins d'approfondissement et de corroboration » (Johnson *et al.*, 2007, p. 123 ; traduction libre). Les méthodes de recherche mixtes sont souvent considérées comme des solutions intermédiaires entre deux positions épistémologiques classiques en recherche : la recherche d'une explication ou d'une vérité universelle et la recherche de vérités multiples pour expliquer un même phénomène. En fait, les méthodes mixtes s'appuient sur une vision pragmatique de la recherche et plusieurs raisons pratiques appuient leur utilisation (Bryman, 2006 ; Greene, Caracelli et Graham, 1989) :

1. la triangulation, pour confirmer ou corroborer une explication ;
2. la complémentarité, pour enrichir, élaborer ou illustrer une explication ;

3. le développement, pour construire de nouveaux devis, choisir des outils de mesure ou un échantillon pour une prochaine expérimentation, etc. ;
4. l'initiation à une nouvelle façon de penser, pour découvrir une nouvelle perspective ou mettre en lumière des paradoxes ou des résultats inattendus, etc. ;
5. l'expansion, pour étendre les possibilités, l'ampleur et la portée d'une expérimentation.

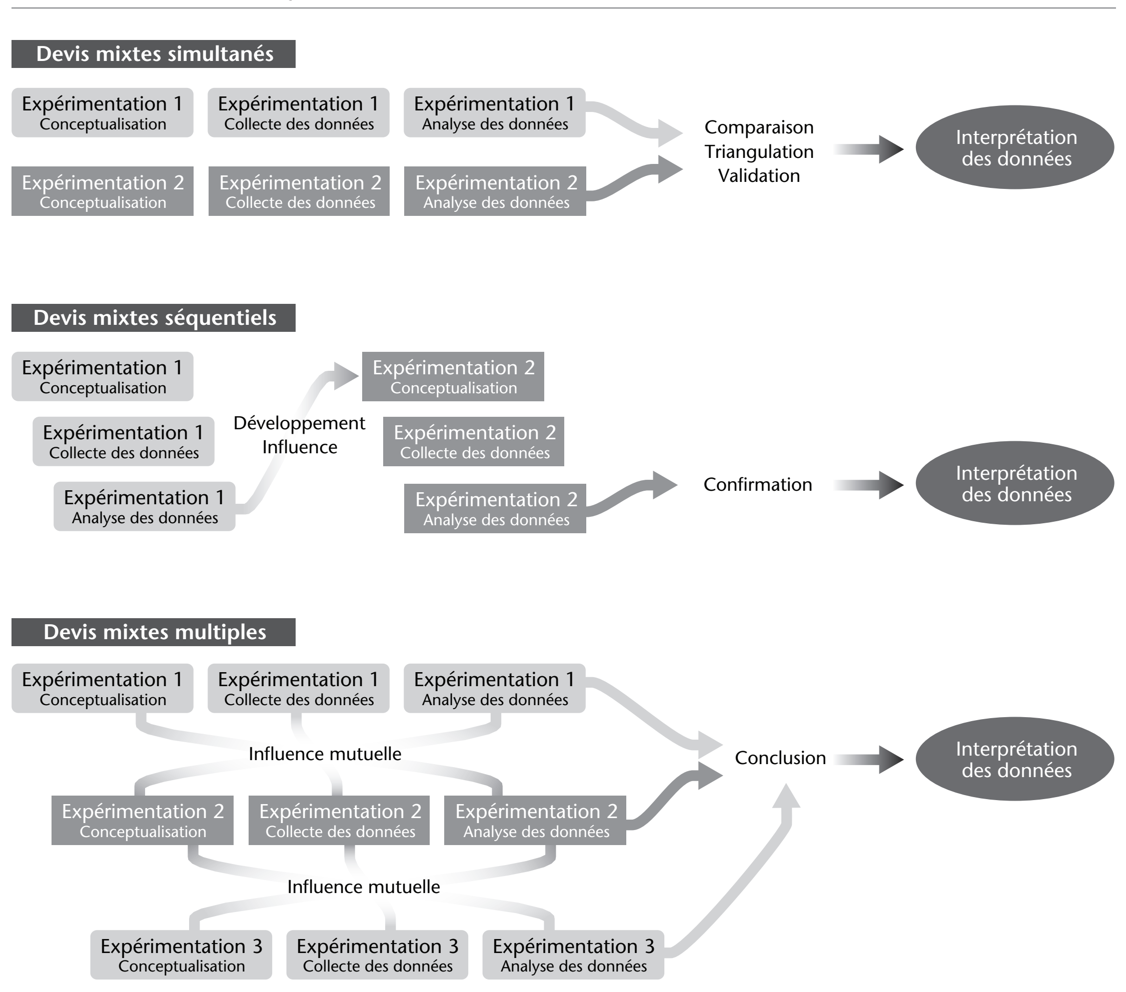
Johnson *et al.* (2007) ont répertorié les définitions de 19 chefs de file dans le domaine des méthodes mixtes et ont relevé cinq dimensions saillantes qui permettent de cerner le type de devis mixte utilisé : 1) Que souhaite-t-on combiner (des devis, des méthodes de collecte de données) ? 2) À quelles étapes du processus de recherche cette combinaison se fait-elle (lors de la collecte de données, lors de l'analyse, etc.) ? 3) À quelle ampleur cette combinaison est-elle présente ? 4) Pour quelles raisons cette combinaison est-elle faite ? 5) Dans quelle orientation de recherche s'inscrit ce choix ?

Au sein de ces cinq dimensions, plusieurs types de devis mixtes peuvent être distingués (Creswell *et al.*, 2003 ; Creswell et Plano Clark, 2011 ; Michaud et Bourgault, 2010 ; Teddlie et Tashakkori, 2009). Tout d'abord, on retrouve les *devis simultanés*, qui sont divisés en quatre types : 1) avec triangulation ; 2) imbriqué ; 3) transformatif ; et 4) avec ou sans transformation de données. Les devis simultanés sont utilisés lorsque les expérimentations quantitatives et qualitatives se déroulent en parallèle et de façon concourante (simultanément) vers un but commun. Ensuite, lorsque les expérimentations quantitatives et qualitatives se déroulent en phases distinctes l'une après l'autre, il existe trois types de *devis séquentiels* : 1) explicatif ; 2) exploratoire ; et 3) transformatif. Enfin, les *devis multiples*, plus récemment apparus dans la littérature des devis mixtes, sont utilisés lorsque des expérimentations simultanées et séquentielles sont présentes dans la même étude selon la phase ou le niveau d'analyse, ou lorsqu'il y a une transformation des données et des influences multiples entre plusieurs expérimentations. La figure 25.1 illustre les trois principaux types de devis mixtes.

Parmi ces trois types de devis, neuf devis spécifiques seront ici présentés. De plus, à la fin de la section 1, le tableau 25.1 présente les différentes appellations des devis mixtes : de la classification classique de Creswell *et al.* (2003) aux classifications plus récentes (Teddlie et Tashakkori, 2009 ; Creswell et Plano Clark, 2011). Ce tableau permet aux lecteurs de bien saisir la terminologie utilisée pour nommer ces différents devis. Aussi, comme références supplémentaires et pour permettre une illustration de

ces neufs devis mixtes, un article servant d'exemple accompagne chacun des devis du tableau 25.1. Chaque article choisi a été publié au cours des dernières années (2008-2013) dans le *Journal of Mixed Methods Research* (<http://mmr.sagepub.com/>) (Mertens, 2011).

Figure 25.1.
Les trois principaux types de devis mixtes



Source : Inspirée de Teddlie et Tashakkori, 2009.

1.1. Devis mixtes simultanés

1.1.1. Devis mixte simultané avec triangulation

Dans le devis mixte simultané avec triangulation, les collectes de données qualitatives et quantitatives se déroulent en parallèle, au sein (ou non) du même échantillon (QUANT et QUAL ou QUAL et QUANT de façon concourante[2]) (Creswell *et al.*, 2003 ; Creswell et Plano Clark, 2011 ; Michaud et Bourgault, 2010 ; Teddlie et Tashakkori, 2009). Les résultats des deux expérimentations sont intégrés à l'étape d'interprétation. Idéalement, les résultats issus de l'expérimentation qualitative et quantitative ont la même priorité, mais en pratique, la priorité peut être accordée à l'une ou à l'autre. Le devis mixte simultané avec triangulation sert à renforcer, confirmer ou corroborer une explication à l'aide de deux expérimentations distinctes au sein d'une même étude (Bryman, 2006 ; Creswell et Plano Clark, 2011 ; Greene *et al.*, 1989). L'avantage d'un devis mixte simultané avec triangulation est que chaque type d'expérimentation est indépendant, c'est-à-dire effectué et analysé séparément (Creswell *et al.*, 2003 ; Creswell et Plano Clark, 2011). Comme les deux expérimentations peuvent se faire en parallèle, l'équipe de recherche peut ainsi gagner du temps. Cependant, un défi relié au devis mixte simultané avec triangulation est la nécessité pour l'équipe de recherche d'être à l'aise avec les deux types de méthode. Aussi, cela peut être un défi de jumeler des résultats qui proviennent de données de différentes formes (chiffrées, textuelles, etc.) ou de combiner des conclusions divergentes. Cela étant dit, il est important dans un devis mixte simultané avec triangulation de bien cibler les données que l'on veut comparer et de prendre le temps de représenter adéquatement les résultats issus de l'expérimentation qualitative (étape de réduction des données), de l'expérimentation quantitative ainsi que ceux jumelés (Creswell et Plano Clark, 2011 ; Teddlie et Tashakkori, 2009).

1.1.2. Devis mixte simultané imbriqué

Dans le devis mixte simultané imbriqué, les collectes de données qualitatives et quantitatives se déroulent en parallèle (qual/QUANT ou quant/QUAL de façon concourante) (Creswell *et al.*, 2003 ; Creswell et Plano Clark, 2011 ; Michaud et Bourgault, 2010 ; Teddlie et Tashakkori, 2009). Cependant, contrairement au devis mixte simultané avec triangulation, une des

2. Les abréviations utilisées font référence aux abréviations proposées par plusieurs auteurs pour illustrer rapidement les différents devis mixtes (Creswell et Plano Clark, 2011 ; Michaud et Bourgault, 2010 ; Teddlie et Tashakkori, 2009) : QUANT/quant = expérimentation quantitative ; QUAL/qual = expérimentation qualitative.

méthodes est prédominante et l'autre y est imbriquée afin d'y documenter une perspective différente ou plus large. Le devis mixte simultané imbriqué permet d'enrichir une explication d'un phénomène (complémentarité), mais aussi d'étendre les possibilités et la portée d'une étude (expansion) (Bryman, 2006 ; Creswell et Plano Clark, 2011 ; Greene *et al.*, 1989). Comme dans le devis mixte simultané avec triangulation, l'avantage d'un devis mixte simultané imbriqué est que chaque type d'expérimentation est indépendant, c'est-à-dire effectué et analysé séparément (Creswell *et al.*, 2003 ; Creswell et Plano Clark, 2011). Contrairement au devis mixte simultané avec triangulation, l'ensemble des membres de l'équipe de recherche n'a pas à maîtriser avec autant d'expertise les deux types de méthodes, l'une prédominant sur l'autre. Cela dit, le défi du devis mixte simultané imbriqué est d'intégrer à l'analyse les données issues de l'expérimentation de moindre priorité. Le danger est qu'elles soient délaissées au profit des données principales, interprétées avec un manque de rigueur ou imbriquées grossièrement. Tout comme dans le devis mixte simultané avec triangulation, l'étape de réduction des données qualitatives est importante afin d'assurer une rigueur à la démarche d'interprétation (Creswell et Plano Clark, 2011 ; Teddlie et Tashakkori, 2009).

1.1.3. Devis mixte simultané transformatif

Tout comme dans les deux devis précédents, le devis mixte simultané transformatif se déroule en deux phases parallèles (QUANT et QUAL) (QUAL et QUANT) (qual/QUANT) (quant/QUAL) (Creswell *et al.*, 2003 ; Creswell et Plano Clark, 2011 ; Michaud et Bourgault, 2010). La principale différence est que le devis mixte simultané transformatif a pour objectif de mener à un changement, à des actions ou de dénoncer des injustices sociales. Le devis mixte simultané transformatif met les méthodes d'analyse qualitative et quantitative au service d'une idéologie de recherche engagée, de recherche-action ou de recherche participative pour redonner du pouvoir aux individus ou aux communautés, et apporter un changement au sein de collectivités marginalisées ou victimes d'injustice sociale (voir le chapitre 26 de cet ouvrage). Le cadre méthodologique utilisé est ainsi appuyé par des modèles théoriques expliquant à la fois le phénomène à l'étude et la démarche de transformation et de changement proposée. L'avantage du devis mixte simultané transformatif est la cohérence du processus méthodologique avec les idéologies et modèles théoriques de l'étude. Enfin, le devis mixte simultané transformatif présente les mêmes avantages et défis que les devis mixtes simultanés avec triangulation ou imbriqué.

1.1.4. Devis mixte simultané avec transformation des données

Dans le devis mixte simultané avec transformation des données, les collectes de données qualitatives et quantitatives se déroulent en parallèle, mais une transformation des données est proposée pour enrichir et poursuivre l'analyse (Creswell et Plano Clark, 2011 ; Teddlie et Tashakkori, 2009). Elle consiste à transformer des données qualitatives en chiffres ou à transformer des données quantitatives en nouvelles variables ou concepts qualitatifs. Par exemple, le chercheur peut être amené à chiffrer des variables issues d'une analyse qualitative pour les introduire dans une analyse de variance (voir le chapitre 16 de cet ouvrage) ou à utiliser les résultats d'une analyse factorielle exploratoire (voir le chapitre 22 de cet ouvrage) pour générer des thèmes qui seront repris dans l'analyse qualitative. Contrairement aux autres devis mixtes simultanés, celui de nature simultanée avec transformation des données utilise généralement le même échantillon pour les deux expérimentations, ce qui n'est pas toujours le cas dans les autres devis mixtes simultanés (Teddlie et Tashakkori, 2009). L'avantage d'un devis mixte simultané avec transformation des données est de pouvoir intégrer aux analyses d'une des deux expérimentations les résultats de l'autre expérimentation. Le défi est cependant de déterminer la meilleure méthode de transformation requise pour les données recueillies et pour répondre aux questions de recherche, ainsi que d'utiliser les données transformées avec rigueur tout au long du processus d'analyse (*inference process* dans Teddlie et Tashakkori, 2009).

1.2. Devis mixtes séquentiels

1.2.1. Devis mixte séquentiel explicatif

Contrairement aux devis mixtes simultanés, le devis mixte séquentiel explicatif se déroule en deux phases chronologiques (QUANT => QUAL) (Creswell *et al.*, 2003 ; Creswell et Plano Clark, 2011 ; Michaud et Bourgault, 2010 ; Teddlie et Tashakkori, 2009). D'abord, il y a une collecte et une analyse de données quantitatives, lesquelles sont prioritaires pour répondre aux questions de recherche. Puis, il y a une phase qualitative où l'équipe de recherche pose de nouvelles questions de recherche pour expliquer les résultats quantitatifs issus de la première expérimentation. Le devis mixte séquentiel explicatif vise à enrichir ou à compléter une explication, ainsi qu'à clarifier les résultats significatifs, non significatifs ou aberrants de la première phase quantitative (Bryman, 2006 ; Creswell et Plano Clark, 2011 ; Greene *et al.*, 1989). Les avantages d'un devis mixte séquentiel explicatif est que l'on procède à une seule collecte de données à la fois, et que chaque

type d'expérimentation est analysé et interprété séparément (Creswell *et al.*, 2003; Creswell et Plano Clark, 2011). Le défi du devis mixte séquentiel explicatif est cependant l'élaboration méthodologique de la seconde phase d'expérimentation (en particulier l'échantillonnage des participants), qui est défini en cours de projet (à la fin de la première phase d'expérimentation) et qui nécessite d'avoir prévu (ou de revérifier) les autorisations éthiques nécessaires (Creswell et Plano Clark, 2011; Teddlie et Tashakkori, 2009). De plus, comme les deux expérimentations se font l'une après l'autre, le devis mixte séquentiel explicatif peut demander beaucoup de temps.

1.2.2. Devis mixte séquentiel exploratoire

Le devis mixte séquentiel exploratoire se déroule également en deux phases chronologiques (QUAL => QUAN), mais cette fois-ci l'ordre des types de méthodes est inversé (Creswell *et al.*, 2003; Creswell et Plano Clark, 2011; Michaud et Bourgault, 2010; Teddlie et Tashakkori, 2009). D'abord, il y a une collecte et une analyse de données qualitatives, lesquelles sont prioritaires pour répondre aux questions de recherche. Puis, il y a une phase quantitative qui permet à l'équipe de recherche de tester ou de généraliser les résultats qualitatifs issus de la première expérimentation. Le devis mixte séquentiel exploratoire vise à confirmer ou à vérifier une explication (complémentarité et triangulation) issue de la première phase qualitative, ou à explorer et concevoir de nouveaux devis quantitatifs lorsque les variables à l'étude sont inconnues, sans modèle théorique ou que le choix des instruments de mesure n'est pas défini (Bryman, 2006; Creswell et Plano Clark, 2011; Greene *et al.*, 1989). Les avantages et les défis du devis mixte séquentiel exploratoire sont les mêmes que le devis séquentiel explicatif (Creswell *et al.*, 2003; Creswell et Plano Clark, 2011). Comme on procède à une seule collecte de données à la fois, on doit définir en cours de projet (ou à la fin de la première phase d'expérimentation) la seconde phase d'expérimentation, et voir au recrutement et aux autorisations éthiques nécessaires (Creswell *et al.*, 2003; Creswell et Plano Clark, 2011). Cette période de préparation vers la seconde phase d'expérimentation peut être très longue étant donné que parfois, la décision de passer à l'élaboration et à la conception d'un nouvel instrument de mesure peut être requise à la suite des variables extraites lors de la première phase qualitative.

1.2.3. Devis mixte séquentiel transformatif

Comme les deux devis précédents, le devis mixte séquentiel transformatif se déroule en deux phases chronologiques (QUAL => QUANT ou QUANT => QUAL) (Creswell *et al.*, 2003; Creswell et Plano Clark, 2011; Michaud et

Bourgault, 2010). La première phase peut être de nature qualitative ou quantitative, suivie de la deuxième phase de nature quantitative ou qualitative, selon la nature de la première phase. La différence avec les deux devis séquentiels précédents est que le devis est transformatif, c'est-à-dire qu'on met les méthodes d'analyse qualitative et quantitative au service d'une idéologie de recherche engagée, de recherche-action ou de recherche participative. Le devis mixte séquentiel transformatif présente les mêmes avantages et pose les mêmes défis que les devis séquentiels (explicatifs ou exploratoires) et le devis simultané transformatif.

1.3. Devis multiples

1.3.1. Devis mixtes multiphases ou multiniveaux

Les devis mixtes multiphases ou multiniveaux combinent les devis mixtes simultanés et les devis mixtes séquentiels en un même devis de recherche (voir la figure 25.1) (Creswell et Plano Clark, 2011; Teddlie et Tashakkori, 2009). Les devis mixtes multiphases ou multiniveaux fournissent un cadre méthodologique global pour un projet incluant plusieurs étapes sur plusieurs années. L'avantage des devis mixtes multiphases ou multiniveaux est de permettre, au sein d'un même devis, la réalisation de différentes étapes en parallèle ou de façon séquentielle et d'intégrer l'influence de ces étapes les unes sur les autres. Le devis mixte multiphases est utilisé dans un contexte d'évaluation de programme où plusieurs phases se succèdent (évaluation des besoins, modélisation logique, évaluation des effets, etc.). Pour ce qui est du devis mixte multiniveaux, il est utilisé dans des contextes d'évaluation de systèmes à différents niveaux d'analyse (locale, régionale, nationale, etc.). Le défi des devis mixtes multiphases ou multiniveaux est la planification et la cohérence de toutes les étapes et de toutes les phases du projet. De plus, il s'agit de devis qui demandent d'importantes ressources et de temps pour bien compléter chaque phase de l'étude tout en assurant une combinaison rigoureuse des méthodes.

1.3.2. Devis mixte complètement intégré

Dans le devis mixte complètement intégré, chaque étape d'une expérimentation influence l'autre (Teddlie et Tashakkori, 2009). La formulation des questions de recherche, les méthodes de collecte de données et l'analyse des données interpellent constamment les deux types d'expérimentation (qualitatives et quantitatives). Les deux méthodes dépendent constamment l'une de l'autre. Le devis mixte complètement intégré partage les avantages

et les défis des autres devis multiples. Il est probablement le plus complexe à mettre en œuvre et il nécessite une expertise dans les deux types de méthodes (quantitatif et qualitatif).

Tableau 25.1.
Différentes appellations des devis mixtes et exemple de publications pour chacun d'eux

Traduction libre des auteurs (inspiré de Michaud et Bourgault, 2010)	Classification classique de Creswell *et al.*, 2003	Autres classifications récentes	Exemple de publications tirées du *Journal of Mixed Methods Research* (2008-2013)
	Les devis mixtes simultanés		
Devis mixte simultané avec triangulation	*Concurrent triangulation design*	▪ *Parrallel mixed design* (Teddlie et Tashakkori, 2009) ▪ *Convergent parallel design* (Creswell et Plano Clark, 2011)	Kramer, J.M. (2011). « Using mixed methods to establish the social validity of a self-report assessment : An illustration using the child occupational self-assessment (COSA) », *Journal of Mixed Methods Research*, vol. 5, nº 1, p. 52-76.
Devis mixte simultané imbriqué	*Concurrent nested design*	▪ *Parrallel mixed design* (Teddlie et Tashakkori, 2009) ▪ *Embedded design* (Creswell et Plano Clark, 2011)	Plano Clark, V.L., K. Schumacher, C. West, J. Edrington, L.B. Dunn, A. Harzstark *et al.* (2013). « Practices for embedding an interpretive qualitative approach within a randomized clinical trial », *Journal of Mixed Methods Research*, vol. 7, nº 3, p. 219-242.
Devis mixte simultané transformatif	*Concurrent transformative design*	▪ *Transformative design* (Creswell et Plano Clark, 2011)	Mertens, D.M. (2012). « Mixed methods as tools for social change », *Journal of Mixed Methods Research*, vol. 5, nº 3, p. 195-197.
Devis mixte simultané avec transformation des données	Non décrit en 2003	▪ *Conversion mixed design* (Teddlie et Tashakkori, 2009) ▪ *Convergent parallel design with data transformation* (Creswell et Plano Clark, 2011)	Jang, E.E., D.E. McDougall, D. Pollon, M. Herbert et P. Russell (2008). « Integrative mixed methods data analytic strategies in research on school success in challenging circumstances », *Journal of Mixed Methods Research*, vol. 2, nº 3, p. 221-247.

(*suite*)

Tableau 25.1. (*suite*)

Les devis mixtes séquentiels			
Devis mixte séquentiel explicatif	*Sequential explanatory design*	▪ *Sequential mixed design* (Teddlie et Tashakkori, 2009) ▪ *Explanatory sequential design* (Creswell et Plano Clark, 2011)	Wesely, P.M. (2010). « Language learning motivation in early adolescents : Using mixed methods research to explore contradiction », *Journal of Mixed Methods Research*, vol. 4, nº 4, p. 295-312.
Devis mixte séquentiel exploratoire	*Sequential exploratory design*	▪ *Sequential mixed design* (Teddlie et Tashakkori, 2009) ▪ *Exploratory sequential design* (Creswell et Plano Clark, 2011)	Palo Stoller, E., N.J. Webster, C.E. Blixen, R.A. McCormick, A.J. Hund, A.T. Perzynski *et al.* (2009). « Alcohol consumption decisions among nonabusing drinkers diagnosed with Hepatitis C : An exploratory sequential mixed methods study », *Journal of Mixed Methods Research*, vol. 3, nº 1, p. 65-86.
Devis mixte séquentiel transformatif	*Sequential transformative design*	▪ *Transformative design* (Creswell et Plano Clark, 2011)	Cambraia Windsor, L. (2013). « Using concept mapping in community-based participatory research : A mixed methods approach », *Journal of Mixed Methods Research*, vol. 7, nº 3, p. 274-293.
Les devis multiples			
Devis mixtes multiphases ou multiniveaux	Non décrit en 2003	▪ *Multilevel mixed design* (Teddlie et Tashakkori, 2009) ▪ *Multiphase design* (Creswell et Plano Clark, 2011)	Ungar, M. et L. Liebenberg (2011). « Assessing resilience across cultures using mixed methods : Construction of the child and youth resilience measure », *Journal of Mixed Methods Research*, vol. 5, nº 2, p. 126-149.
Devis mixte complètement intégré	Non décrit en 2003	▪ *Fully integrated mixed design* (Teddlie et Tashakkori, 2009)	Brown, R.A., D.P. Kennedy, J.S. Tucker, D. Golinelli et S.L. Wenzel (2013). « Monogamy on the street : A mixed methods study of homeless men », *Journal of Mixed Methods Research*, vol. 7, nº 4, p. 328-346.

1.4. Critères de qualité des devis mixtes

Que l'on fasse le choix d'un devis simultané, séquentiel ou multiple, les devis mixtes doivent répondre aux mêmes règles de qualité que tout autre processus de recherche. Chaque expérimentation, qu'elle soit qualitative ou quantitative, doit respecter les critères de qualité d'une recherche scientifique : crédibilité/validité interne, transférabilité/validité externe, fiabilité/fidélité, neutralité/objectivité (Teddlie et Tashakkori, 2009). Creswell et Plano Clark (2011) proposent différentes stratégies pour minimiser les biais possibles associés aux méthodes mixtes : 1) prendre au sein de la même population à l'étude les échantillons pour les deux expérimentations (pour les devis mixtes séquentiels, on peut utiliser le même échantillon d'individus) ; 2) proposer une seule question de recherche qui pourra être répondue par les deux expérimentations ou impliquant les deux types d'expérimentation ; 3) utiliser des procédures de transformation les plus simples possibles, c'est-à-dire qui demeurent près des résultats primaires, qui impliquent des analyses statistiques simples (p. ex., fréquence, pourcentage), ou qui permettent de vérifier la validité des conclusions ; 4) dans le cas de résultats contradictoires, réexaminer les données, poursuivre la collecte de données, ouvrir les conclusions à des nouvelles hypothèses si l'explication ne se trouve pas dans les limites méthodologiques ; 5) tenter de présenter avec la même attention les résultats des deux expérimentations (sauf pour le devis mixte simultané imbriqué) ; 6) utiliser un modèle théorique pour relier les différentes étapes, niveaux et expérimentations.

Lorsqu'une méthode de recherche mixte est utilisée, l'équipe de chercheurs doit porter une attention particulière à la façon d'analyser ensemble les données, à la manière de présenter les données issues des différents types d'expérimentation et à la qualité des inférences (Franz, Worrell et Vögele, 2013). L'utilisation de figures, de descriptions claires des procédures de collecte de données, de tableaux de comparaison côte à côte ou croisés par thème, de transformations de données sont quelques exemples de stratégies de présentation des devis mixtes et de comparaison de données suggérées par Creswell et Plano Clark (2011, chapitre 7) et Teddlie et Tashakkori (2009, chapitre 12). D'autres stratégies et recommandations sont également présentées dans l'article de Franz *et al.* (2013) sur l'intégration des données, telles que l'utilisation d'un cadre de référence pour circonscrire les étapes et le processus du devis mixte.

1.5. Publication des résultats de projets de recherche avec un devis mixte

La publication des résultats d'études intégrant des devis mixtes peut poser à certains égards un défi. La multitude des données et analyses produites ainsi que la complexité des étapes et procédures peuvent parfois être difficiles à présenter brièvement et simplement dans le cadre d'un même article scientifique. Creswell et Plano Clark (2011) proposent une structure de rédaction et d'évaluation des devis mixtes, en soulignant l'importance de détailler la méthodologie utilisée et les raisons qui expliquent le choix de ce type de devis. Les règles de publication du *Journal of Mixed Methods Research* (<http://mmr.sagepub.com/>) (Mertens, 2011) indiquent d'ailleurs l'importance d'intégrer explicitement les données des deux types d'expérimentation au sein du même article, ainsi que de relever la contribution et les enjeux du devis mixte. À noter que le *Journal of Mixed Methods Research* permet des soumissions de 8 000 à 10 000 mots, soit plus du double des mots acceptés par les autres revues dans le domaine de la santé mentale et les domaines connexes.

2. ILLUSTRATION : ANALYSE DES EFFETS CLINIQUES ET FONCTIONNELS D'UN HÔPITAL DE JOUR PSYCHIATRIQUE

L'étude choisie aux fins de l'illustration est une évaluation des effets cliniques, fonctionnels et économiques d'un hôpital de jour psychiatrique, en comparaison à l'hospitalisation (Larivière *et al.*, 2010a, 2010b; Larivière et Boyer, 2011). Les hôpitaux de jour sont des services en santé mentale qui ont été développés au cours des dernières décennies comme alternative moins coûteuse à l'hospitalisation pour traiter des personnes avec des troubles mentaux en décompensation aiguë et qui sont moins à risque de régression et de dépendance (Casarino, Wilner et Maxer, 1982). Toutefois, les effets cliniques et fonctionnels des hôpitaux de jour par rapport à l'hospitalisation n'étaient pas clairement démontrés avec profondeur, ni le profil des personnes qui pouvaient en profiter davantage. Ces raisons ont mené à la réalisation de l'étude, qui a intégré un *devis mixte séquentiel explicatif* pour répondre à ses objectifs. Comme décrit dans la première partie de ce chapitre, ce type de devis débute par une collecte et une analyse de données quantitatives, suivies par une phase qualitative pour expliquer les résultats quantitatifs (objectifs de triangulation ou de complémentarité).

Dans cette étude, les objectifs principaux étaient de comparer les effets cliniques et la satisfaction à l'égard d'un hôpital de jour psychiatrique pour des personnes adultes et âgées présentant divers troubles mentaux aigus (groupe expérimental) avec ceux de l'hospitalisation (groupe de comparaison). Dans un premier temps, un devis quasi expérimental, avec

un groupe de comparaison non aléatoire et non équivalent, a été utilisé (Contandriopoulos *et al.*, 1990). Les trois temps de mesure étaient : 1) au début du suivi dans les services index (T1) ; 2) cinq à huit semaines plus tard (T2) ; et 3) six mois après le T2 (T3). Pour cerner plus en profondeur les effets cliniques et fonctionnels de l'hôpital de jour psychiatrique, trois types de clientèles ont été comparés : 1) des personnes avec un trouble psychotique ; 2) des personnes avec un trouble de l'humeur et anxieux ; 3) des personnes avec un trouble de la personnalité du groupe B (p. ex., trouble de la personnalité limite). Dans un deuxième temps, l'équipe de recherche avait comme objectif d'explorer de façon qualitative l'expérience d'un suivi en hôpital de jour psychiatrique et ses effets dans la vie des personnes six mois suivant le T2 (soit le T3). Pour ce faire, un dispositif qualitatif suivant principalement les critères de Creswell (2007) et de Poupart *et al.* (1997) a été proposé pour l'échantillonnage et la procédure de collecte de données (voir le chapitre 1 de cet ouvrage).

2.1. Méthodes

2.1.1. Phase quantitative

De façon plus précise, pour la phase quantitative, l'équipe de recherche a suivi les standards reconnus pour mener cette étape. Par exemple, le calcul de la taille de l'échantillon a été effectué avec des paramètres précis (p. ex., différence standardisée modérée [*effect size*] de 0,52 et 20 % de participants supplémentaires en prévision des pertes au suivi). Les variables dépendantes ont été mesurées avec des questionnaires quantitatifs validés tels que le Symptom-Checklist-90-Revised (Derogatis, 1994 ; Fortin et Coutu-Walkulczyk, 1985) et la Mesure des habitudes de vie (version 3.1 abrégée) (Fougeyrollas *et al.*, 2002). Les analyses quantitatives ont été réalisées avec le logiciel SPSS. Les caractéristiques sociodémographiques des participants ont été documentées avec des statistiques descriptives, telles que des moyennes et des pourcentages. Pour comparer les changements aux trois temps de mesure entre les deux groupes (hôpital de jour et hospitalisation), des analyses de variance à mesures répétées ont été effectuées.

2.1.2. Phase qualitative

Pour la phase qualitative, un échantillon par contraste (Poupart *et al.*, 1997), avec des critères d'inclusion théorique (Strauss et Corbin, 1990) et provenant du groupe *hôpital de jour* (groupe expérimental), a été formé. Ce choix de type d'échantillon permettait de construire une mosaïque d'expériences

à partir de cas diversifiés. Le contraste a été établi par l'intermédiaire de variables stratégiques : 1) des variables générales comme le sexe et l'âge et 2) des variables spécifiques rattachées directement au problème étudié comme l'équipe clinique et le degré de satisfaction des services au congé (selon la cote au questionnaire Verona Service Satisfaction Scale [VSSS-EU] rempli au T2). Le choix de ces variables découlait d'études antérieures ou d'hypothèses théoriques qui justifiaient leur importance en tant que sources probables de différence (Poupart *et al.*, 1997). Ainsi, les participants provenaient des six équipes cliniques de l'hôpital de jour (équipes divisées selon le diagnostic principal et l'âge) et ont été regroupés en trois catégories : 1) sous la moyenne au VSSS-EU par rapport à l'ensemble des participants (insatisfaits) ; 2) dans la moyenne au VSSS-EU par rapport à l'ensemble des participants (satisfaits) ; et 3) au-dessus de la moyenne au VSSS-EU par rapport à l'ensemble des participants (très satisfaits).

Avant de commencer la collecte de données, l'équipe de recherche a estimé qu'un participant par niveau de satisfaction et par équipe clinique (n = 18) permettrait d'atteindre la saturation des données. Les participants ont été sélectionnés au fur et à mesure du déroulement de la collecte de données. Des entrevues individuelles semi-dirigées ont été réalisées six mois après le congé de l'hôpital de jour. Les objectifs de l'entrevue étaient d'explorer le vécu des participants durant la période de six mois suivant le congé de l'hôpital de jour et de refaire une évaluation des services de l'hôpital de jour avec un recul de six mois. Le guide d'entrevue a été élaboré par la chercheuse principale, une clinicienne de l'hôpital de jour et la chef clinico-administrative de l'hôpital de jour. Il a été validé par trois experts pour vérifier la clarté des questions et s'assurer de la séquence de ces dernières. Un prétest a été effectué auprès de deux personnes suivies à l'hôpital de jour. L'entrevue finale était divisée en quatre thèmes : 1) changements perçus ; 2) apprentissages faits ; 3) ingrédients thérapeutiques de l'hôpital de jour ; 4) suggestions pour améliorer les services reçus. Les entrevues ont été transcrites intégralement et vérifiées par la chercheuse principale.

Les recommandations de Miles et Huberman (1994) ont été suivies lors de l'analyse qualitative du contenu des entrevues. La première étape de l'analyse a été la codification du contenu écrit. La codification correspond à un processus d'attribution d'une unité de sens, nommée « codes », à des mots, des phrases ou des paragraphes (voir le chapitre 1 de cet ouvrage). Pour définir les codes, un guide a été conçu par l'équipe de recherche. Ce guide incluait les thèmes, les catégories et les sous-catégories, leurs définitions et des citations de participants en guise d'exemples. Tout au long de l'analyse, la chercheuse principale a utilisé un journal de mémos afin de noter ses réflexions sur les concepts, les liens entre ceux-ci, des synthèses sur les

thèmes communs et ses questionnements. Afin de déterminer une catégorie ou un thème, des dictionnaires et des écrits scientifiques ont été consultés. De plus, pour augmenter la validité de l'analyse, une clinicienne avec de l'expérience en hôpital de jour psychiatrique et en recherche qualitative a codé en entier une des premières entrevues, sélectionnée au hasard. Une version bonifiée du guide de codes a été validée par un membre de l'équipe de recherche lorsque 80 % du matériel a été analysé et que le guide de codes était relativement stable. La fidélité code-recode a également été vérifiée à deux moments au cours de l'analyse, avec deux semaines d'intervalle entre chaque codification. Par ailleurs, les commentaires partagés spontanément par 15 autres participants ont été utilisés comme source de validation des thèmes, des catégories et des sous-catégories. Ils ont aussi permis de confirmer l'atteinte de la saturation des idées. L'analyse qualitative a été structurée à l'aide du logiciel informatique QDA Miner (version 3.0). Les analyses quantitatives et qualitatives ont été effectuées séparément.

2.2. Résultats

2.2.1. Résultats des analyses quantitatives

Aux fins de l'illustration, une partie des résultats des analyses des effets cliniques et fonctionnels sont présentés[3]. Les analyses quantitatives ont montré que pour des personnes avec des troubles de l'humeur et anxieux ou des troubles psychotiques en phase aiguë[4], un suivi en hôpital de jour démontrait des effets supérieurs à l'hospitalisation en ce qui concerne la gravité des symptômes, l'estime de soi, la réalisation de la participation sociale globale et la satisfaction face à la participation sociale, particulièrement dans les rôles sociaux. À la suite d'un suivi à l'hôpital de jour, ces mêmes participants ont continué à démontrer une amélioration significative dans leur estime de soi, leur réalisation globale de leur participation sociale et leur satisfaction face à leur réalisation d'activités courantes et de rôles sociaux. De leur côté, les personnes hospitalisées sont demeurées stables sur toutes les variables. Toutefois, après le congé des deux services, le degré de changement était similaire entre les deux groupes (l'ensemble de ces résultats sont détaillés dans les tableaux 25.2 et 25.3). Il faut souligner qu'une analyse d'équivalence des deux groupes a été effectuée et qu'il n'y avait aucune différence statistiquement significative sur l'ensemble des

3. Le lecteur intéressé par les autres résultats (dont le volet économique) sont invités à lire les articles publiés par Larivière *et al.*, 2010a ; 2010b.
4. Les personnes avec des troubles de la personnalité du groupe B n'ont pas été incluses dans ces analyses, car il n'y en avait pas dans le groupe hospitalisation.

variables sociodémographiques et cliniques examinées au T1 (c.-à-d. âge, sexe, état civil, niveau de scolarité, nombre de conditions psychiatriques comorbides, nombre d'épisodes antérieurs, coté à l'Échelle d'évaluation globale du fonctionnement et variables dans le tableau 25.2).

Tableau 25.2.
Comparaison de l'évolution clinique entre les deux groupes (T1-T2)

Variables	Hôpital de jour (*n* = 40)			Hospitalisation (*n* = 13)			Interaction[7] *p*
	Moyenne (é.t.) T1	Moyenne (é.t.) T2	T2-T1[6] *p*	Moyenne (é.t.) T1	Moyenne (é.t.) T2	T2-T1[6] *p*	
Gravité de la psychopathologie (participants)[1]	1,33 (0,68)	0,75 (0,50)	< 0,001	1,24 (0,64)	0,97 (0,60)	0,22	0,047
Gravité de la psychopathologie (professionnel)[2]	4,58 (2,19)	3,00 (2,23)	< 0,001	4,23 (2,58)	2,93 (2,94)	0,12	0,82
Estime de soi[3]	27,10 (6,52)	31,33 (5,31)	< 0,001	28,54 (7,79)	27,92 (7,24)	0,67	0,01
Participation sociale – réalisation (total)[4]	6,84 (1,21)	7,50 (0,97)	< 0,001	6,90 (1,16)	7,00 (1,17)	0,46	0,02
Participation sociale – satisfaction (total)[5]	2,72 (0,70)	3,16 (0,49)	0,001	2,64 (0,70)	2,74 (0,73)	0,78	0,008

[1] SCL-90-R : 1 = pas du tout ; 4 = extrêmement.
[2] SCL-90 Analogue : 1 = pas du tout et 10 = extrêmement.
[3] Échelle d'estime de soi de Rosenberg : cote plus élevée = meilleure estime de soi.
[4] MHAVIE : 0 = non réalisé et 9 = réalisé sans difficulté et sans aide.
[5] MHAVIE : 1 = très insatisfait et 4 = très satisfait.
[6] Test Wilcoxon's Signed Ranks.
[7] Test Mann Whitney U.

Tableau 25.3.
Comparaison de l'évolution clinique entre les deux groupes (T2-T3)

Variables	Hôpital de jour (*n* = 35)			Hospitalisation (*n* = 10)			Interaction[2] *p*
	Moyenne (é.t.) T2	Moyenne (é.t.) T3	T3-T2[1] *p*	Moyenne (é.t.) T2	Moyenne (é.t.) T3	T3-T2[1] *p*	
Gravité de la psychopathologie (participants)	0,77 (0,50)	0,66 (0,61)	0,14	0,97 (0,66)	0,95 (0,62)	0,39	0,72
Gravité de la psychopathologie (professionnel)	3,38 (2,60)	4,29 (2,54)	0,21	3,75 (3,23)	3,99 (2,59)	0,74	0,64
Estime de soi	31,29 (5,37)	33,23 (5,07)	0,04	27,90 (7,64)	26,40 (6,24)	0.55	0,15
Participation sociale – réalisation (total)	7,49 (0,92)	7,85 (0,73)	0,02	6,97 (1,31)	7,18 (1,46)	0,29	0,97
Participation sociale – satisfaction (total)	3,13 (0,48)	3,33 (0,55)	0,07	2,85 (0,78)	2,82 (0,73)	0,72	0,88

[1] Test Wicoxon's Signed Ranks.
[2] Test Mann Whitney U.

Les analyses quantitatives ont également fait ressortir que durant le suivi à l'hôpital de jour, les personnes avec des troubles psychotiques se sont moins améliorées que les personnes avec des troubles de l'humeur et anxieux et de la personnalité par rapport à la gravité des symptômes, la détresse et l'estime de soi. Les trois groupes se sont améliorés de façon similaire par rapport à leur participation sociale globale. Par ailleurs, alors que les personnes avec des troubles psychotiques et de la personnalité du groupe B sont demeurées stables après le congé de l'hôpital de jour, les personnes avec un trouble de l'humeur et anxieux ont continué à s'améliorer de façon significative par rapport à leur estime de soi et à leur participation sociale.

2.2.2. Résultats des analyses qualitatives

Les entretiens qualitatifs ont amené spontanément les 18 participants à tracer un bilan de leur parcours avant, pendant et après le suivi à l'hôpital de jour. Celui-ci a permis de valider les effets quantitatifs de l'hôpital de

jour et d'ajouter à leur compréhension, ainsi que de proposer des nouveaux éclairages. Le résumé des thèmes découlant des analyses qualitatives est présenté dans le tableau 25.4.

Tableau 25.4.
Résumé des thèmes découlant des analyses qualitatives

Avant le début du suivi à l'hôpital de jour	▪ Réticences initiales à venir à l'hôpital de jour ▪ Délai d'attente de deux semaines et plus : période difficile
Pendant l'hôpital de jour : *Qu'est-ce que l'expérience a représenté pour les participants ?*	▪ Expérience aidante et transformatrice ▪ A provoqué des changements marquants ▪ Tremplin pour poursuivre le cheminement
Pendant l'hôpital de jour : *Qu'est-ce que les participants retiennent de leur séjour ?*	▪ Contact interpersonnel avec tous ▪ Qualités des intervenants ▪ Thérapie individuelle et de groupe ▪ Activités thérapeutiques ▪ Intensité du suivi
Après l'hôpital de jour : *Comment se sont déroulés les six premiers mois ?*	▪ La « perte » de l'hôpital de jour laisse un vide ▪ À travers les hauts et les bas, le cheminement vers le bien-être mental continue ▪ Le passage à l'action : une clé pour une transition positive
Services et interventions à maintenir à l'hôpital de jour (quelques exemples)	▪ Combinaison thérapie de groupe et individuelle, avec approches et modalités variées ▪ Équipes multidisciplinaires ▪ Continuer d'offrir des repas à coûts réduits, des places de stationnement gratuites et le service de taxi à coût modique
Services et interventions à améliorer à l'hôpital de jour (quelques exemples)	▪ Mieux organiser concrètement le suivi en externe pendant le séjour ▪ Poursuivre le suivi sous une forme moins intensive jusqu'à l'amorce concrète du prochain suivi en externe (p. ex., groupe de soutien) ▪ Inviter d'anciens participants de l'hôpital de jour à témoigner à propos de leur cheminement

2.3. Intégration des résultats quantitatifs et qualitatifs

L'étape suivante est l'intégration de l'ensemble des résultats pour en dégager une synthèse. Cette intégration a clairement mis en lumière le fait que les résultats qualitatifs ont apporté plusieurs types d'enrichissement aux résultats quantitatifs. En premier lieu, ils ont permis de comprendre certains résultats quantitatifs plus en profondeur (objectif de complémentarité). Par

exemple, concernant l'évolution stable entre les T2 et T3 pour les trois clientèles diagnostiques de l'hôpital de jour sur différentes variables cliniques et fonctionnelles, deux des thèmes principaux touchant la période posthôpital de jour étaient « À travers les hauts et les bas, le cheminement continue » et « Le passage à l'action est une clé pour une transition positive ». Ces thèmes complètent les cotes obtenues en montrant que cette période fut un processus non linéaire et que des stratégies furent mises en œuvre durant cette période. De plus, un élément plus approfondi dans les entrevues soutenant cette stabilité était les diverses formes de soutien social reçus par les participants de la part de leur entourage. Dans le même sens, parmi les changements positifs perçus par les participants, certains n'étaient pas toujours des éléments directement mesurés dans les questionnaires (p. ex., de meilleures habiletés relationnelles, plus d'affirmation de soi ou utiliser la résolution de problèmes au quotidien). Ceci s'applique aussi à certaines difficultés persistantes soulevées (p. ex., des ajustements avec les effets de la médication ou ressentir un certain isolement). Par ailleurs, avec six mois de recul, les participants ont exprimé à nouveau leur appréciation de la formule de l'hôpital de jour, dont les qualités humaines et professionnelles des intervenants, la combinaison des interventions de groupe et individuelles, et les modalités thérapeutiques proposées. Ces résultats sont congruents avec les cotes moyennes au questionnaire sur la satisfaction à l'égard des services reçus (VSSS-EU) rempli au T2 (tableau 25.5). Dans les entrevues, les participants ont toutefois détaillé des modalités précises appréciées telles que les activités projectives et l'établissement d'un objectif hebdomadaire. Ils ont également discuté de certaines réserves par rapport à la thérapie de groupes lorsque celle-ci intégrait plus de dix participants et ont proposé des suggestions d'améliorations concrètes.

Aussi, les analyses qualitatives ont permis d'apporter de nouveaux résultats (objectif d'expansion). Par exemple, l'un des thèmes majeurs dégagés par rapport à la période suivant la fin de l'hôpital de jour était la transition initiale dans les premières semaines. Cette fin de suivi a été perçue par plusieurs participants comme une « *cassure* », une « *coupure drastique* », comme « *se retrouver dans le désert* ». Plusieurs ont soulevé divers problèmes d'accès à la thérapie à la suite du congé de l'hôpital de jour, corroborés en partie par nos données recueillies sur l'utilisation des services, et ont précisé davantage ces obstacles. Par exemple, certains psychiatres ne donnaient pas suite aux recommandations de l'hôpital de jour ou aux demandes des participants ou encore, les services avaient un horaire mal adapté aux participants qui travaillent durant la journée à temps complet. Toutes les suggestions d'amélioration et les stratégies pour favoriser une transition positive après le congé de l'hôpital de jour faisaient également partie des nouveaux éclairages apportés par les données qualitatives.

Tableau 25.5.
Satisfaction à l'égard de l'hôpital de jour (*n* = 60)

1 = très mauvaise **5 = excellente**	**Moyenne (é.t.)**	**Étendue**
Total	4,34 (0,39)	3,28-4,90
Satisfaction générale	4,64 (0,46)	3,33-5,00
Comportements des professionnels	4,62 (0,42)	3,04-5,00
Information	4,10 (0,68)	2,67-5,00
Accès	4,37 (0,59)	3,00-5,00
Efficacité	4,49 (0,52)	3,13-5,00
Types d'interventions	3,98 (0,38)	3,00-4,65
Implication de la famille	4,00 (0,94)	1,00-5,00

En somme, lorsque les résultats des deux types d'analyses ont été combinés, il a été conclu qu'un hôpital de jour psychiatrique organisé en équipes cliniques avec des clientèles homogènes et proposant un suivi thérapeutique intégrant diverses modalités d'interventions en groupe et en individuel est un service apprécié. De plus, il a des effets bénéfiques chez des personnes avec des troubles psychotiques, des troubles de l'humeur et anxieux, et des troubles de la personnalité du groupe B, et ce, jusqu'à six mois après le congé. Malgré le fait que 98 % des participants de l'hôpital de jour ont consulté des services de santé mentale au cours des six premiers mois suivant leur congé de ce service, d'après leurs témoignages, il est clair que la continuité du suivi après l'hôpital de jour est un aspect à améliorer, en particulier l'accès à la thérapie non médicale. Pour favoriser une transition positive, plusieurs stratégies aidantes apprises à l'hôpital de jour ont été déployées par les participants, telles que continuer à se fixer des objectifs, s'auto-observer au quotidien et remettre dans sa vie des activités significatives (projets, intérêts, loisirs).

CONCLUSION

Les questions de recherche nécessitant plusieurs types de méthodes et de sources de données sont nombreuses dans le domaine de la santé mentale et des sciences sociales. Les méthodes de recherche mixtes répondent à ce besoin et proposent différents types de devis qui intègrent des éléments des méthodes qualitatives et quantitatives à l'intérieur d'un même projet.

Toute la richesse obtenue d'un processus de recherche mixte pose cependant certains défis. Le chercheur et l'équipe de recherche doivent détenir les compétences nécessaires permettant l'utilisation rigoureuse de plusieurs types de méthodes et leur intégration. Les étapes et les procédures peuvent être multiples et exiger des échéanciers plus longs pour assurer un processus rigoureux et une interinfluence judicieuse des deux types de méthodes. La présentation et la publication des résultats peuvent nécessiter de la créativité et de prendre le temps pour bien illustrer la contribution des devis mixtes. Ces contributions sont d'ailleurs nombreuses. Les devis mixtes permettent de trianguler des informations provenant de différentes sources de données, d'élaborer davantage une explication d'un phénomène, de construire de nouveaux devis et outils de mesure, de découvrir de nouvelles perspectives, ainsi que d'étendre la portée d'une étude au-delà des devis habituels. Comme tous les types de devis de recherche, les méthodes de recherche mixtes doivent répondre à de hauts standards de qualité scientifique.

RÉFÉRENCES

Brown, R.A., D.P. Kennedy, J.S. Tucker, D. Golinelli et S.L. Wenzel (2013). « Monogamy on the street : A mixed methods study of homeless men », *Journal of Mixed Methods Research*, vol. 7, nº 4, p. 328-346.

Bryman, A. (2006). « Integrating quantitative and qualitative research : How is it done ? », *Qualitative Research*, vol. 6, nº 1, p. 97-113.

Cambraia Windsor, L. (2013). « Using concept mapping in community-based participatory research : A mixed methods approach », *Journal of Mixed Methods Research*, vol. 7, nº 3, p. 274-293.

Casarino, J.P., M. Wilner et J.T. Maxey (1982). « American Association for Partial Hospitalization (AAPH) standards and guidelines for partial hospitalization », *International Journal of Partial Hospitalization*, vol. 1, p. 5-21.

Contandriopoulos, A.-P., F. Champagne, L. Potvin, J.-L. Denis et P. Boyle (1990). *Savoir préparer une recherche. La définir, la structurer, la financer*, Montréal, Les Presses de l'Université de Montréal.

Creswell, J.W. (2007). *Qualitative Inquiry et Research Design. Choosing Among Five Approaches*, Thousand Oaks, Sage Publications.

Creswell, J.W. et V.L. Plano Clark (2011). *Designing and Conducting Mixed Methods Research*, 2e éd., Thousand Oaks, Sage Publications.

Creswell, J.W., V.L. Plano Clark, M.L. Gutmann et W.E. Hanson (2003). « Advanced Mixed methods research design », dans A. Tashakkori et C. Teddlie (dir.), *Handbook of Mixed Methods in Social and Behavioral Research*, Thousand Oaks, Sage Publications, p. 209-240.

Derogatis, L.R. (1994). *SCL-90-R. Symptom-Checklist-90-R. Administration, Scoring and Procedures Manual*, Minneapolis, National Computer Systems.

FORTIN, M.F. (2010). *Fondements et étapes du processus de recherche. Méthodes quantitatives et qualitatives*, Montréal, Chenelière Éducation.

FORTIN, M.F. et G. COUTU-WALKULCZYK (dir.) (1985). *Validation et normalisation d'une mesure de santé mentale : le SCL-90-R*, Rapport présenté au Conseil québécois de la recherche sociale (CQRS), Montréal, Faculté des sciences infirmières, Université de Montréal.

FOUGEYROLLAS, P., L. NOREAU, S.-A. DION *et al*. (dir.) (2002). *La mesure des habitudes de vie* (version 3.1), Lac-Saint-Charles, Réseau international sur le processus de production du handicap.

FRANZ, A., M. WORRELL et C. VÖGELE (2013). « Integrating mixed method data in psychological research : Combining Q methodology and questionnaires in a study investigating cultural and psychological influences on adolescent sexual behavior », *Journal of Mixed Methods Research*, vol. 7, nº 4, p. 370-389.

GREENE, J.C., V.J. CARACELLI et W.F. GRAHAM (1989). « Toward a conceptual framework for mixed-method evaluation designs », *Education Evaluation and Policy Analysis*, vol. 11, p. 255-274.

JANG, E.E., D.E. MCDOUGALL, D. POLLON, M. HERBERT et P. RUSSELL (2008). « Integrative mixed methods data analytic strategies in research on school success in challenging circumstances », *Journal of Mixed Methods Research*, vol. 2, nº 3, p. 221-247.

JOHNSON, R.B., A.J. ONWUEGBUZIE et L.A. TURNER (2007). « Toward a definition of mixed methods research », *Journal of Mixed Methods Research*, vol. 1, nº 2, p. 112-133.

KRAMER, J.M. (2011). « Using mixed methods to establish the social validity of a self-report assessment : An illustration using the child occupational self-assessment (COSA) », *Journal of Mixed Methods Research*, vol. 5, nº 1, p. 52-76.

LARIVIÈRE, N. et R. BOYER (2011). « Revisiter l'expérience vécue en hôpital de jour six mois après le congé : parcours et perceptions des personnes utilisatrices », *Santé mentale au Québec*, vol. 36, nº 1, p. 75-97.

LARIVIÈRE, N., J. DESROSIERS, M. TOUSIGNANT et R. BOYER (2010a). « Revisiting the day hospital experience six months after discharge : How was the transition and what have clients retained ? », *Psychiatric Quarterly*, vol. 81, p. 81-96.

LARIVIÈRE, N., J. DESROSIERS, M. TOUSIGNANT et R. BOYER (2010b). « Who benefits the most from psychiatric day hospitals ? A comparison between three clinical groups », *Journal of Psychiatric Practice*, vol. 16, nº 2, p. 93-102.

MERTENS, D.M. (2011). « Publishing mixed methods research », *Journal of Mixed Methods Research*, vol. 5, nº 1, p. 3-6.

MERTENS, D.M. (2012). « Mixed methods as tools for social change », *Journal of Mixed Methods Research*, vol. 5, nº 3, p. 195-197.

MICHAUD, C. et P. BOURGAULT (2010). « Les devis de recherche non traditionnels », dans M.F. Fortin (dir.), *Fondements et étapes du processus de recherche. Méthodes quantitatives et qualitatives*, 2e éd., Montréal, Chenelière Éducation, p. 656.

MILES, M.B. et A.M. HUBERMAN (1994). *Qualitative Data Analysis*, 2e éd., Thousand Oaks, Sage Publications.

PALO STOLLER, E., N.J. WEBSTER, C.E. BLIXEN, R.A. MCCORMICK, A.J. HUND, A.T. PERZYNSKI *et al.* (2009). « Alcohol consumption decisions among nonabusing drinkers diagnosed with Hepatitis C : An exploratory sequential mixed methods study », *Journal of Mixed Methods Research*, vol. 3, nº 1, p. 65-86.

PLANO CLARK, V.L., K. SCHUMACHER, C. WEST, J. EDRINGTON, L.B. DUNN, A. HARZSTARK *et al.* (2013). « Practices for embedding an interpretive qualitative approach within a randomized clinical trial », *Journal of Mixed Methods Research*, vol. 7, nº 3, p. 219-242.

POUPART, J., J.-P. DESLAURIERS, L.-H. GROULX, A. LAPERRIÈRE, R. MAYER et A.P. PIRES (1997). *La recherche qualitative : enjeux épistémologiques et méthodologiques*, Boucherville, Gaëtan Morin.

STRAUSS, A. et J. CORBIN (1990). *Basics of Qualitative Research. Grounded Theory Procedures and Techniques*, Newbury Park, Sage Publications.

TEDDLIE, C. et A. TASHAKKORI (2009). *Foundations of Mixed Methods Research : Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*, Thousand Oaks, Sage Publications.

UNGAR, M. et L. LIEBENBERG (2011). « Assessing resilience across cultures using mixed methods : Construction of the child and youth resilience measure », *Journal of Mixed Methods Research*, vol. 5, nº 2, p. 126-149.

WESELY, P.M. (2010). « Language learning motivation in early adolescents : Using mixed methods research to explore contradiction », *Journal of Mixed Methods Research*, vol. 4, nº 4, p. 295-312.

CHAPITRE 26

LES APPROCHES DE RECHERCHE PARTICIPATIVES
Illustration d'un partenariat pour l'amélioration des pratiques de réadaptation en santé mentale au Québec

Nadine Larivière
Catherine Briand
Marc Corbière

FORCES

- Même si certaines approches ne visent pas directement un changement de pratique, elles tiennent compte des acteurs du milieu et de leur évolution, et favorisent ainsi l'appropriation et l'application des résultats par les acteurs concernés.
- Elles permettent de donner une voix à des personnes qui traditionnellement ne participent pas à toutes les étapes d'une recherche (p. ex., des utilisateurs de services, des gestionnaires).
- Elles facilitent le décloisonnement entre la recherche et la pratique.

LIMITES

- Maintenir l'intérêt et la participation de tous les acteurs tout au long du projet peut être un défi pour le chercheur.
- Les projets de recherche intégrant une approche participative peuvent prendre beaucoup de temps avant d'arriver à terme.
- Il existe un certain flou concernant l'application de cette approche, notamment lorsque la recherche est sujette à des aléas sur lesquels les chercheurs n'ont pas le contrôle.

Depuis plusieurs décennies et dans plusieurs domaines, dont les sciences de la santé (Abma, 2000 ; Cockburn et Trentham, 2002 ; Mercier, 1997), l'évaluation de programmes (Dussault *et al.*, 2012 ; Ridde, 2009), les sciences sociales (Anderson-Draper, 2006 ; Healy, 2001) et l'éducation (Cousins et Walker, 2000 ; Desgagné, 2007 ; Morrissette, 2012), des auteurs réfléchissent et documentent un type d'approche où les chercheurs, les praticiens, les gestionnaires, les utilisateurs de services ou les proches collaborent à la réalisation d'études émanant de préoccupations du terrain. Ces approches de recherche ont plusieurs appellations selon la discipline à l'origine de leur développement, mais elles ont toutes une trame transversale. Dans les écrits, on retrouve les termes *évaluation participative*, *évaluation centrée sur l'empowerment*, *évaluation démocratique*, *évaluation développementale*, *recherche-action*, *recherche collaborative*, *recherche-formation* ou *recherche-intervention* (Anadón et Couture, 2007 ; Cousins et Chouinard, 2012).

Le principe fondamental des approches participatives est l'engagement actif de toutes les parties prenantes ou détenteurs d'enjeux[1] dans la prise de décision concernant le processus de recherche. Ce processus requiert la mise en commun continuelle des connaissances, des savoir-faire et des ressources, ainsi que le partage du pouvoir entre les participants (Simard *et al.*, 1997).

La plupart des définitions des approches de recherche participative comprennent les trois éléments suivants : 1) la *recherche* : la collecte systématique d'informations en fonction de la question de recherche et des besoins des personnes concernées ; 2) l'éducation : l'occasion d'apprentissage pour tous les participants et l'amélioration des habiletés individuelles et collectives ; et 3) l'*action* : la mise en œuvre d'une action concertée (Simard *et al.*, 1997). Toutes les approches participatives permettent d'établir un lien entre la recherche et l'action, entre la théorie et la pratique, entre la logique du chercheur et celle des détenteurs d'enjeux ou encore, de coconstruire le savoir (Anadón et Couture, 2007). À la suite de ce premier survol des ingrédients essentiels des approches participatives, le présent chapitre situera, dans une première partie, le développement historique de celles-ci dans divers domaines, définira les types d'approches participatives courants, et décrira ses particularités et ses enjeux. Ensuite, dans la seconde partie, un projet de recherche évaluative participative réunissant différents détenteurs

1. Plusieurs termes sont utilisés pour désigner les participants du milieu qui sont concernés par l'objet de l'étude (p. ex., un programme ou un problème commun à un groupe). Notamment, on retrouve : les « parties prenantes », les « détenteurs d'enjeux », les « acteurs du milieu ou du terrain », les « acteurs locaux » ou les « utilisateurs de connaissances », souvent désignés par le terme anglais « *stakeholders* ».

d'enjeux impliqués dans l'offre de services de réadaptation en santé mentale au Québec servira d'illustration afin de mieux comprendre l'application de cette approche.

1. BREF HISTORIQUE DES APPROCHES PARTICIPATIVES

En recherche sociale, la recherche participative a pris racine dans les approches de développement communautaire et dans l'activisme pour la reconnaissance des droits humains. Ainsi, outre par le partage de la propriété du contenu des projets de recherche, la recherche participative se caractérise par le fait que les problèmes sociaux sont analysés sous un angle communautaire et orientés vers une action également communautaire (Kemmis et McTaggart, 2008).

En évaluation de programmes, ces mouvements sociaux ont aussi eu une influence, et ce, surtout à partir des années 1970. Plusieurs chercheurs dans ce domaine ont souligné l'importance de prendre en compte la culture des programmes, de situer les évaluations dans leur contexte social et de répondre aux besoins manifestés dans ces contextes (Brisolura, 1998). Notant que les résultats des évaluations étaient souvent sous-utilisés, Michael Quinn Patton et d'autres chefs de file en évaluation de programmes ont dû réfléchir sur des manières de favoriser l'appropriation et l'application des résultats (Patton, 1997). Par ailleurs, des valeurs de prises de décision démocratique, qui impliquait les détenteurs d'enjeux dans la démarche évaluative, ont graduellement été intégrées en évaluation de programmes, tout comme les nouveaux rôles de négociateur et de facilitateur chez le chercheur-évaluateur (Brisolura, 1998). En 1989, Guba et Lincoln ont proposé une quatrième génération d'évaluation pour dépasser les limites des générations précédentes. Dans ces dernières, les évaluations de programmes étaient surtout centrées sur les besoins des décideurs et des gestionnaires pour aider à la prise de décision sur l'allocation des ressources. Elles visaient principalement à améliorer l'efficacité des programmes et adoptaient surtout le paradigme positiviste, ce qui impliquait, entre autres, que les chercheurs-évaluateurs devaient être externes au programme évalué pour demeurer neutres et objectifs. Guba et Lincoln (1989) ont fait ressortir que les éléments contextuels, politiques et culturels étaient à prendre en compte dans l'évaluation, stipulant ainsi qu'il était nécessaire d'impliquer tous les acteurs concernés par le programme en évaluation et non seulement les gestionnaires et décideurs (p. ex., clinicien, utilisateur de service, proche). Cousins et Withmore (1998) ont dégagé deux grands courants en évaluation participative qui s'appliquent encore actuellement : 1) l'évaluation participative pratique ; et 2) l'évaluation participative émancipatrice (*transformative evaluation*) (Cousins et Chouinard, 2012 ; Ridde, 2006).

L'évaluation participative pratique a pris forme aux États-Unis et au Canada. Sa fonction principale est de soutenir l'utilisation de l'évaluation, avec la prémisse implicite que l'évaluation est dirigée vers un programme, une politique ou une prise de décision organisationnelle. L'hypothèse au cœur de ce type d'évaluation est que la participation des détenteurs d'enjeux va permettre d'augmenter la pertinence de l'évaluation, l'appropriation et par le fait même l'utilisation du processus et des résultats (Cousins et Withmore, 1998). Dans ce type d'évaluation participative, les membres de la communauté de l'évaluation travaillent en partenariat égalitaire avec les membres de la communauté du programme pour implanter des évaluations qui, typiquement, visent à soutenir des initiatives d'améliorations de programmes. Les participants sont impliqués dans une variété de tâches liées à la recherche, dont la détermination des questions et des objectifs de la recherche, l'analyse des données et la diffusion des résultats (Weaver et Cousins, 2005). Un des exemples connus de l'évaluation participative pratique est l'évaluation centrée sur les résultats de Patton (Patton, 2008 ; Patton et LaBossière, 2012).

L'évaluation participative transformatrice, quant à elle, intègre des principes de participation et d'action afin de démocratiser le changement social. Elle est née au début des années 1970 dans des pays comme l'Inde et la Colombie en réaction aux modèles positivistes de recherche qui étaient perçus par certains chercheurs comme étant abusifs et détachés des problèmes économiques et sociaux urgents (Cousins et Chouinard, 2012). De plus, ce mouvement affirmait qu'il fallait se préoccuper des populations les plus pauvres, laissées pour compte, et que la recherche devait créer des liens avec des actions sociopolitiques développées par des organisations sociales (Anadón et Savoie-Zajc, 2007 ; Fals Borda, 1985). Les travaux de Paolo Freire (1973) ont permis d'établir les fondements philosophiques de ce type d'évaluation (Cousins et Withmore, 1998). Brièvement, pour ce dernier, il importait d'assurer la participation des secteurs populaires de la société et de transformer leur réalité. De plus, pour Freire, la production et la communication des connaissances sont deux processus simultanés (Anadón et Savoie-Zajc, 2007). Il est important de noter ici que dans d'autres ouvrages (Anadón et Savoie-Zajc, 2007), les travaux et la pensée de Freire et de Fals Borda sont associés au développement de ce que l'on nomme la recherche-action. L'un des exemples connus de l'évaluation participative transformatrice est l'évaluation de quatrième génération de Guba et Lincoln (1989).

1.1. Types d'approches participatives

L'histoire montre que selon le domaine où l'approche participative a été conçue, appliquée et sa finalité, le terme pour la nommer et la définir varie. La section qui suit décrira les principaux types d'approches participatives que l'on retrouve en évaluation de programmes et en sciences sociales, soit l'*évaluation participative*, la *recherche-action participative* et la *recherche collaborative*.

En évaluation de programmes, la mise en œuvre d'un processus évaluatif (plutôt qu'uniquement des méthodes ou des techniques spécifiques), selon une démarche de partenariat entre des évaluateurs et des personnes qui sont concernées par l'évaluation, mais qui ne sont pas des évaluateurs, se nomme évaluation participative (Cousins et Chouinard, 2012 ; Ridde, 2006). Daigneault et Jacob (2012) indiquent que ce qui distingue une évaluation participative d'une évaluation non participative se retrouve dans les trois dimensions essentielles suivantes : le contrôle du processus évaluatif (évaluateur ou parties prenantes), la sélection des parties prenantes (étendue des types d'acteurs qui sont inclus) et l'étendue de l'implication des parties prenantes.

Pour accompagner cette réflexion, Daigneault, Jacob et Tremblay (2012) ont conçu un outil avec ces trois attributs clés, qui permet de déterminer le degré de participation des parties prenantes[2]. Pour les trois attributs, un score à cinq niveaux allant de 0 = implication ou contrôle nul des parties prenantes/contrôle total de l'évaluateur à 1 = implication ou contrôle total des participants/contrôle nul de l'évaluateur peut être attribué. Tout d'abord, concernant l'étendue de l'implication des parties prenantes dans les quatre grandes étapes de l'évaluation (conception, collecte et analyse, formulation des jugements et recommandations et diffusion des résultats), plus le nombre de tâches dans lesquelles interviennent les parties prenantes (excluant le chercheur-évaluateur) est élevé, plus l'évaluation est participative.

Il faut l'implication d'au moins un type d'acteurs (ou parties prenantes) pour que l'évaluation soit considérée comme participative. Si l'évaluateur dirige complètement la démarche et que les participants ne font qu'observer ou fournir des données, l'évaluation n'est pas considérée comme participative. Il a été ainsi établi que pour considérer une évaluation comme participative, il faut un score minimal de 0,25 pour chaque attribut, ce qui correspond à une implication d'un type de parties prenantes dans au moins une étape de l'évaluation avec un contrôle limité (le contrôle est surtout

2. Les lecteurs intéressés à connaître davantage l'outil sont invités à lire les textes suivants de Daigneault, Jacob et Tremblay, 2012, et de Daigneault et Jacob, 2012.

de la part de l'évaluateur). Pour compléter cette évaluation du degré de participation, deux autres dimensions pertinentes peuvent aussi être examinées, soit la facilité de la mise en œuvre de l'évaluation (de fluide/gérable à rigide/ingérable) et les rapports de pouvoir entre les parties prenantes (de conflictuels à neutres) (Daigneault et Jacob, 2012; Ridde, 2006; Weaver et Cousins, 2005). D'autres outils existent comme les *Guidelines and Categories for Classifying Participatory Research Projects*, qui incluent une perspective de promotion de la santé à propos de ces questions (Green *et al.*, 1995).

En sciences sociales, la *recherche-action participative* est une démarche à long terme, qui vise à résoudre des problèmes de groupe tels que ceux liés à l'inégalité ou à l'injustice sociale, et ce, en remettant en question les pratiques ou les structures oppressives, puis en les changeant (Anadón et Savoie-Zajc, 2007; Michaud et Bourgault, 2010). Elle vise également à améliorer les conditions de vie d'une communauté (Anadón et Savoie-Zajc, 2007). Elle est considérée comme un outil intellectuel au service de la population et facilite la prise de position des acteurs dans l'espace public (Anadón et Savoie-Zajc, 2007). Idéalement, le chercheur fait partie du groupe et tous les détenteurs d'enjeux (cochercheurs) sont considérés comme égaux, sans hiérarchie (Anadón et Savoie-Zajc, 2007; Michaud et Bourgault, 2010). Les détenteurs d'enjeux doivent accepter de jouer un rôle actif dans le processus de changement (Michaud et Bourgault, 2010). La démarche de recherche repose sur un système de cycles en trois phases: 1) l'analyse de la situation et du changement à réaliser (recherche); 2) l'analyse conjointe des résultats, de la rétroaction et de la formation; et 3) la réflexion sur les résultats de la collecte de données, la planification et la mise en œuvre du changement à accomplir (action). Ce plan inclut des objectifs, des tâches, des responsables, un échéancier, des ressources et des indicateurs de l'atteinte. Ces phases sont parfois reprises sur plusieurs cycles (Michaud et Bourgault, 2010). Les chercheurs utilisant la recherche-action participative mettent en place plusieurs mécanismes d'*empowerment*, de collaboration et de prises de décision, comme le développement des capacités, l'autogestion, la formation de comités de citoyens ou des ateliers de discussion (Anadón et Savoie-Zajc, 2007). La recherche- action participative permet de produire trois types de savoirs: 1) des savoirs académiques transférables; 2) des savoirs de pratiques; et 3) des savoirs d'expérience. Ces derniers sont issus des réflexions du chercheur et des participants pour dégager comment cette équipe de recherche a su élaborer et mettre en œuvre sa capacité de résolution de problèmes et de collaboration, et ce, pour le mieux-être d'une communauté (Guillemette, 2011).

En éducation, la *recherche collaborative* est définie comme une démarche de coconstruction de savoirs entre chercheurs et praticiens (Desgagné, 2007), où l'objet d'étude est le savoir d'action du praticien. Elle consiste en deux volets : 1) une activité formelle de recherche qui s'intéresse à conceptualiser les savoirs issus de la pratique et 2) l'offre aux praticiens d'activités réflexives susceptibles de leur être utiles pour leur développement professionnel (Morrissette, 2012). Les praticiens qui s'impliquent dans cette démarche ne sont pas simplement des informateurs pour les chercheurs, mais doivent s'engager à explorer (un aspect de) leur pratique et à livrer à l'équipe leur compréhension contextuelle (Desgagné et Bednarz, 2005). Les activités réflexives doivent conduire à une reconstruction de récits de pratiques qui pourraient être disponibles par la suite pour la formation de futurs enseignants et qui pourraient aussi permettre une théorisation de récits prototypes (Desgagné, 2007). Dans ce type de recherche, le chercheur est préoccupé à la fois par le volet recherche et par le volet formation. Il doit permettre un métissage des logiques et des sensibilités pratiques et théoriques (Desgagné, 2007). Le chercheur doit clairement établir que le but n'est pas d'améliorer directement les pratiques ou d'offrir une formation avec du contenu à transmettre, mais bien de réfléchir, de faire le point et d'expliciter les expériences de pratique significatives (Desgagné, 2007 ; Morrissette, 2012).

Il existe une variante en gestion nommée *investigation collaborative et développementale* proposée par Baron (2007), qui s'est inspirée de Heron et Reason (1997), ainsi que de Torbert (1999). Elle vise à la fois la recherche *et* le développement professionnel. Elle part d'une communauté de pratique[3] de six à douze personnes qui, dans le cas de Heron et Reason (1997), articulent un projet commun puis commencent plusieurs cycles d'action et de réflexion, soutenus par les chercheurs.Dans le cas de Torbert (1999), chaque membre de la communauté de pratique formule un projet personnel visant à dépasser certaines difficultés récurrentes rencontrées dans son expérience de gestion. Chaque participant est invité à explorer des incidents critiques, à les revisiter pour faire des nouvelles prises de conscience, à faire de la résolution de problèmes en groupe et ensuite, à expérimenter les nouvelles avenues qui sont ressorties des rencontres.

L'examen de ces définitions et contextes d'application montrent que les approches de recherche participatives peuvent servir à diverses fins telles que trouver des solutions à des problèmes de santé et des moyens pour mettre en œuvre ces solutions (Simard *et al.*, 1997 ; Ridde, 2006), évaluer

3. Les termes *communauté de pratique* signifient « un groupe de personnes ayant en commun un domaine d'expertise ou une pratique professionnelle, et qui se rencontrent pour échanger, partager et apprendre les uns des autres, en face à face ou virtuellement » (Tremblay, 2005, p. 695).

les effets d'un programme ou la qualité des services (Simard *et al.*, 1997; Ridde, 2006), soutenir le développement professionnel et la construction d'un savoir pertinent lié à la pratique (Desgagné, 2007) ou conceptualiser des théories (Morrissette, 2012). Avoir recours à une approche de recherche participative plutôt qu'à une autre peut alors se justifier par les trois types de raisons suivantes: 1) des raisons pragmatiques: accroître les chances de réussite du processus de recherche et de l'utilisation des résultats; 2) des raisons politiques: dans un souci démocratique, de participation citoyenne et d'émancipation; et 3) des raisons épistémologiques ou philosophiques: la pluralité des points de vue est essentielle à la construction d'un argument, car il n'existe pas de réalité unique (Cousins et Chouinard, 2012; Weaver et Cousins, 2005).

Afin de mener une recherche participative, certaines grandes étapes sont importantes. Les étapes proposées dans la figure 26.1 sont inspirées des travaux des divers auteurs cités dans ce chapitre.

1.2. Enjeux des approches participatives pour les chercheurs

Le chercheur qui décide d'utiliser une approche de recherche participative verra rapidement la possibilité d'adopter plusieurs rôles qui pourront varier selon la finalité et les étapes de la recherche: organisateur, responsable de la démarche de coconstruction et du processus de recherche permettant d'en dégager le sens, vulgarisateur de son expertise et de ses connaissances, animateur, guide, médiateur (Cousins et Chouinard, 2012; Couture, Bednarz et Barry, 2007; Savoie-Zajc, 2012). De plus, dès le début de la recherche, le chercheur doit clarifier les rôles et les responsabilités de chacun, incluant le sien. Cette clarification des rôles doit aussi être validée pendant la recherche, car les intentions des acteurs peuvent évoluer, ce qui peut créer de nouveaux rapports entre eux (Savoie-Zajc, 2012). Une suggestion proposée par Savoie-Zajc (2012) dans le cadre de la recherche-action, mais qui peut s'appliquer à toutes les approches de recherche participative, est de former un comité de pilotage composé de représentants des divers acteurs pour anticiper les réactions des personnes aux diverses étapes et proposer des façons pour inciter les personnes à s'engager et à collaborer. Le partage des responsabilités est donc une question au cœur de ces approches.

Figure 26.1.
Étapes clés d'une recherche participative

ÉTAPE 1 – Première rencontre entre le chercheur et certains acteurs du milieu pour :

a) traduire les préoccupations en questions de recherche ;

b) planifier les types de parties prenantes à impliquer dans le processus ainsi que les individus potentiels à retenir ;

c) comprendre le contexte et les enjeux associés (p. ex., la pression organisationnelle pour démontrer la pertinence d'un programme afin d'assurer sa pérennité

ÉTAPE 2 – Rencontres avec les représentants des parties prenantes pour :

a) définir le projet ;

b) clarifier les rôles, attentes et fonctionnement de l'équipe ;

c) établir les étapes (de la conception à la diffusion des résultats/transfert des connaissances) et les responsabilités de chacun dans celles-ci ainsi que l'estimation de l'échéancier (en fonction des fonds de recherche et du temps disponible des acteurs du milieu) ;

d) anticiper les obstacles à la participation et à la réalisation du projet et proposer des solutions.

ÉTAPE 3 – Collecte et analyse des données : devis et méthodes qualitatives, quantitatives ou mixtes. La validation des résultats est très importante pour que les parties prenantes puissent les confirmer, les clarifier ou proposer des suggestions.

ÉTAPE 4 – Planification de la diffusion des résultats :

a) types de diffusion :
- communications écrites (p. ex., articles scientifiques ou professionnels, blogue)
- communications orales : conférence, atelier, affiche ;

b) tâches et les responsabilités des parties (incluant les chercheurs) par rapport à ces communications ;

c) auditoires visés pour la diffusion (p. ex., intervenants et gestionnaires d'autres établissements avec le même type de programme, décideurs).

Tout au long des étapes ci-dessus, le chercheur devra :

1. fournir des formations vulgarisées (p. ex., concepts/modèles théoriques, méthodes de recherche, questionnaires, logiciels, pratiques pour analyses qualitatives avec une partie d'une entrevue) ;
2. assurer une validation des rôles et attentes;
3. maintenir une atmosphère de confiance (p. ex., donner une voix à tous ; discuter des malaises observés ou explicites);
4. soutenir la motivation de l'équipe (p. ex., rencontres bilans; présentation des résultats préliminaires) ;
5. assurer une validation des résultats.

Selon les domaines, des différences sont observées. Par exemple, en gestion, selon l'approche de Heron et Reason (1997), toutes les responsabilités des chercheurs sont partagées entre les participants, et ce, à toutes les étapes de la recherche. Dans l'évaluation dite de quatrième génération, il y a un partage de la reddition de comptes et de la coconstruction, mais l'évaluateur garde le contrôle sur le déroulement technique de l'évaluation pour mieux jouer ses rôles de catalyseur et d'agent de changement (Guba et Lincoln, 1989; Stufflebeam, 2001). Dans cette approche, l'évaluateur adopte une position indépendante des détenteurs d'enjeux pour rester ouvert aux différentes conceptions de la réalité véhiculées par les acteurs en présence et non pour contrôler des biais potentiels (Guba et Lincoln, 1989).

Les enjeux de pouvoir sont également inévitables en recherche participative, et le chercheur doit prendre le temps de bien comprendre le contexte et la culture du ou des milieux impliqués (Cousins et Chouinard, 2012), de déterminer à qui revient la responsabilité de choisir les participants (chercheur? demandeur de l'évaluation?) et de réfléchir aux statuts des divers participants dans l'équipe (Savoie-Zajc, 2012). Ces réflexions ont des implications éthiques et, par conséquent, le chercheur doit toujours garder à l'esprit comment il prend ces statuts en compte pour éviter que les rencontres ne deviennent des occasions de compétition ou d'intimidation qui inhibent la participation des membres de l'équipe. Si le chercheur provient de l'interne de l'établissement ou de l'organisation à l'étude et qu'il veut réaliser une recherche participative de type recherche-action, cet enjeu sera d'autant plus important que le chercheur aura à faire face à la méfiance éventuelle de collègues qui pourraient lui prêter des objectifs cachés (Savoie-Zajc, 2012). Ainsi, les personnes impliquées dans la sélection des participants représentant les parties prenantes sont tout aussi importantes que les personnes retenues pour la recherche elle-même (Cousins et Chouinard, 2012). De plus, le chercheur doit prendre le temps de bien sélectionner le terme pour nommer les parties prenantes: seront-ils des «cochercheurs», des «partenaires», des «coproducteurs de savoir», des «facilitateurs pour donner accès à des participants» (p. ex., personnes de leur propre réseau; utilisateurs de services)? (Couture *et al.*, 2007; Savoie-Zajc, 2012). Le choix des termes donnés aux acteurs ou leur appellation a des conséquences sur la propriété des données, du travail et de la diffusion des résultats (Savoie-Zajc, 2012). Tous ces aspects décrits dans cette section montrent bien toute l'importance accordée au caractère relationnel lors d'une démarche de recherche participative.

Dans toutes les approches participatives, la collecte de données peut se faire avec des méthodes quantitatives, qualitatives ou mixtes, selon la question de recherche. Diverses sources de données peuvent servir à la collecte

de données telles que l'analyse documentaire, les entrevues individuelles ou de groupe (pour les entrevues de groupes, voir le chapitre 12 du présent ouvrage), des questionnaires, des grilles, des sondages, de l'observation participante ou non participante, ou des journaux de bord (Michaud et Bourgault, 2010; Ridde, 2006). Peu importe le type d'approche participative, le chercheur se doit d'être flexible pour ajuster parfois ses méthodes de collecte de données et prendre le temps de s'assurer que tous les acteurs de la recherche les connaissent bien. Ceci implique que le chercheur doive prendre du temps pour offrir une formation et de l'accompagnement (*coaching*) afin d'augmenter le degré de participation des parties prenantes (Cousins et Chouinard, 2012).

Malgré les divers défis présentés jusqu'à présent, une approche de recherche participative bien menée peut générer plusieurs retombées positives pour les parties prenantes sur le plan individuel, interpersonnel et organisationnel. Une recension de 121 études participatives (Cousins et Chouinard, 2012) a fait ressortir que par exemple, sur le plan personnel, il y avait plus de réflexivité chez les cochercheurs, plus d'ouverture à se diriger vers de nouvelles orientations ou une augmentation de la confiance en soi et de ses capacités. Sur le plan organisationnel, des changements positifs dans la culture du milieu ont été notés, dont l'intégration de la recherche évaluative.

1.3. Critères de scientificité

Couture, Bednarz et Barry (2007) ont proposé quelques critères de scientificité propres aux approches participatives. Tout d'abord, il importe de s'interroger sur la pertinence sociale des connaissances produites, incluant l'immédiateté de l'action et la synchronie entre la production et le transfert. De plus, la congruence est nécessaire entre les observations du chercheur et les perceptions des parties prenantes au processus, ou entre l'expérience et la compréhension de la pratique. Enfin, la «double vraisemblance» du processus et des résultats pour la communauté professionnelle et la communauté scientifique est de mise. La «double vraisemblance», telle que décrite par Dubet (1994), implique que le chercheur doive élaborer un devis et des méthodes de recherche reconnues et comprises, lesquelles permettront aux parties prenantes d'atteindre la finalité visée, qui peut être de divers ordres (p. ex., le changement social, une réflexion sur l'objet de préoccupation mutuelle, le développement professionnel ou l'amélioration d'un programme). Il est fortement suggéré de commenter ces types de critères dans une publication portant sur une recherche participative.

En conclusion, jumeler la recherche avec les préoccupations issues des milieux de la pratique ou de la communauté et impliquer les parties prenantes dans le processus de production de nouvelles connaissances procurent de nombreux avantages. Toutefois, la recherche participative comporte ses enjeux particuliers, qui demandent au chercheur de les anticiper, de les préparer et de faire preuve d'une souplesse pour s'ajuster tout au long de la démarche de recherche.

2. ILLUSTRATION D'UN PARTENARIAT POUR L'AMÉLIORATION DES PRATIQUES DE RÉADAPTATION EN SANTÉ MENTALE AU QUÉBEC

2.1. But et contexte du projet

Le projet de recherche qui sert d'illustration à ce chapitre est une recherche évaluative participative qui emprunte à la fois aux principes de l'évaluation participative et à ceux de la recherche-action, définies dans la première partie de ce chapitre. Quoique ce projet tente de transformer les mécanismes de prise de décision par l'implication d'un plus grand nombre de détenteurs d'enjeux dans l'amélioration des services de réadaptation en santé mentale, il a une visée principalement pragmatique centrée sur l'utilisation des données d'évaluation. Plus précisément, ce projet a comme but principal d'évaluer les services de réadaptation en santé mentale offerts dans trois régions types du Québec et d'établir, par un processus démocratique et itératif entre différents détenteurs d'enjeux, un plan d'action vers l'amélioration des services offerts.

Ce projet a débuté lors d'une rencontre de réseautage sur le thème *Partenariat pour l'amélioration des pratiques de réadaptation au Québec visant le rétablissement des personnes atteintes de troubles mentaux graves* qui s'est tenue à Montréal en mai 2009 (grâce à une subvention de partenariat). Cette rencontre a réuni : quatorze gestionnaires responsables de programmes de santé mentale adulte au sein de centres de santé et de services sociaux (CSSS) de douze régions administratives du Québec ; deux chefs de services et un conseiller clinique au sein de centres spécialisés de santé mentale adulte (Institut universitaire en santé mentale de Québec, Institut universitaire en santé mentale Douglas, Centre de réadaptation la Myriade) ; quatre représentants d'associations provinciales en santé mentale concernées par le sujet (Association québécoise de la réadaptation psychosociale – AQRP, Association des personnes utilisatrices de services en santé mentale – APUR, Société québécoise de la schizophrénie – SQS) ; et six membres de l'équipe

de recherche (trois chercheurs, deux assistants de recherche et un décideur directeur de services de santé mentale à l'Institut universitaire en santé mentale Douglas, qui agissait à titre de cochercheur dans le projet).

Lors de cette rencontre de deux journées complètes, l'ensemble des participants ($n = 27$), soutenus et animés par l'équipe de recherche, ont eu à : 1) partager à tour de rôle leur réalité respective (organisation des services et enjeux principaux) ; 2) discuter d'une définition opérationnelle et concrète de la réadaptation et du rétablissement ; 3) partager sur les pratiques exemplaires de réadaptation actuellement offertes au Québec et celles à valoriser davantage auprès des personnes vivant avec des troubles mentaux graves (p. ex., schizophrénie) ; et 4) déterminer en sous-groupes des moyens pour mieux intégrer ces pratiques exemplaires à l'offre de services et faciliter leur diffusion entre les régions du Québec. L'une des propositions émises lors de ces journées a alors été de soumettre conjointement un projet de recherche qui permettrait d'approfondir dans une perspective locale la compréhension des enjeux qui limitaient à ce moment l'implantation des meilleures pratiques de réadaptation dans les régions du Québec.

2.2. Méthodes

En collaboration avec les partenaires présents lors de ces deux journées, trois régions types du Québec ont été choisies pour faire partie du projet d'évaluation : 1) la région de l'Est de l'Île de Montréal (496 918 habitants, soit le quart de la région de Montréal) (zone urbaine de 4620,7 hab/km^2) ; 2) la région de Lanaudière (476 941 habitants) (zone mixte ou semi-urbaine de 36,8 hab/km^2) ; 3) la région de l'Abitibi-Témiscamingue (146 753 habitants) (zone rurale de 2,5 hab/km^2). Ces régions ont été choisies non seulement parce qu'elles représentaient des contextes différents, mais aussi parce que les partenaires de ces régions étaient disponibles pour s'investir dans le projet. Aussi, ces trois régions servaient de régions témoins pour généraliser la réflexion à l'ensemble des régions du Québec et pour l'alimenter.

Pour ce projet d'une durée de trois ans, le devis utilisé a été une étude de cas multiples (Yin, 2014) où chaque cas est une des trois régions à l'étude qui a la responsabilité d'offrir à sa population ayant un trouble mental grave des services qui répondent à l'ensemble des besoins de réadaptation (services de réadaptation de proximité, spécialisés et surspécialisés) (voir aussi le chapitre 4 de cet ouvrage). Chaque région représente en quelque sorte un système organisé d'actions (tel que défini par Contandriopoulos et Souteyrand, 1996, et Contandriopoulos 2003) qui a pour mission d'offrir une gamme de services de réadaptation à sa population aux prises avec un trouble mental grave. L'évaluation des services de réadaptation de chacune

des régions avait pour objectifs de : 1) décrire de façon détaillée la programmation de réadaptation offerte dans chacune des régions à l'étude ; 2) définir les écarts et les enjeux qui nuisent à l'implantation des meilleures pratiques de réadaptation dans l'offre de services des trois régions ciblées ; 3) établir les priorités et actions à entreprendre pour soutenir les intervenants et les gestionnaires dans l'amélioration de l'offre de services (sur les plans local et au niveau provincial pour l'ensemble des régions du Québec).

Ces objectifs ont été formulés par l'équipe de recherche à la suite de discussions lors des deux journées de réseautage et en collaboration avec les partenaires de chaque région (grâce à une subvention de planification[4]).

Lors de chaque étape du projet, le processus d'évaluation s'est fait en étroite collaboration avec les détenteurs d'enjeux de chacune des trois régions. Ces détenteurs d'enjeux ont été proposés par des partenaires qui avaient participé aux deux journées de réseautage et qui avaient été sélectionnés lors des rencontres de planification. Une attention particulière a été accordée à la représentation de l'ensemble des parties prenantes : des représentants gestionnaires de chaque organisation impliquée dans l'offre de services de réadaptation de la région (à la fois des organisations publiques et communautaires), des représentants d'utilisateurs de services et de proches, des représentants de l'Agence de la santé et des services sociaux ou d'un organisme décideur. Ainsi, pour chacune des trois régions, un comité consultatif regroupant 14 à 17 détenteurs d'enjeux a été mis sur pied. Ces comités consultatifs ont été impliqués lors de chaque étape du projet, à raison de trois à cinq rencontres par année pendant toute la durée du projet. Tous les détenteurs d'enjeux ont été impliqués dans les trois objectifs susmentionnés : 1) la description et la validation de la programmation de réadaptation (l'offre de services), 2) la définition et la validation des écarts et des enjeux qui nuisent à l'implantation des meilleures pratiques de réadaptation, 3) le choix des pistes d'action et des priorités. Par ailleurs, ils ont aussi participé à certaines décisions concernant les méthodes d'évaluation, notamment la rédaction des canevas d'entrevue et des histoires de cas. Selon les critères de Daigneault et Jacob (2012), le degré de participation des parties prenantes dans ce projet correspond à une cote de 1,75 sur 3 (implication modérée aux étapes du projet 0,5/1, diversité substantielle des parties prenantes impliquées 0,75/1, contrôle égal entre l'évaluateur et les parties prenantes au processus évaluatif 0,5/1).

4. Il est à noter que pour toute la durée du projet et pour soutenir chaque étape en étroit partenariat, l'équipe de recherche a reçu quatre subventions différentes : une subvention de partenariat pour organiser les deux journées de réseautage, une subvention de planification pour commencer le travail étroit avec les trois régions choisies, une subvention de fonctionnement pour exécuter le projet planifié ainsi qu'une subvention de dissémination.

Pour atteindre les objectifs de l'étude, plusieurs étapes ont été réalisées, et pour la majorité d'entre elles, cela s'est faiten étroite collaboration avec les comités consultatifs. Le tableau 26.1 résume ces étapes.

Tableau 26.1.
Étapes réalisées dans le projet illustré

PLANIFICATION	
Étape 1	Établir un contact avec les différentes régions et sélectionner les détenteurs d'enjeux.
Étape 2	Mettre en place les comités consultatifs.
Étape 3	Discuter et convenir de la démarche et des méthodes d'évaluation auprès des comités consultatifs de chaque région (incluant le partage des rôles et des tâches).
ÉVALUATION	
Étape 4	Rédiger les canevas d'entrevue et les histoires de cas en étroite collaboration avec les comités consultatifs de chaque région (processus itératif entre les trois régions et l'équipe de recherche).
Étape 5	Établir une liste d'intervenants terrain pour participer aux entrevues.
Pour assurer la confidentialité des intervenants terrain qui ont participé aux entrevues, une liste d'intervenants était proposée par le comité consultatif, mais le choix final a été fait par l'équipe de recherche.	
Étape 6	Construire les modèles logiques[1] de la programmation de réadaptation offerte dans chacune des régions à partir des documents remis sur les services offerts (rapports annuels, sites Web, dépliants d'information, etc.).
Étape 7	Schématiser et valider les modèles logiques de la programmation de réadaptation par l'entremise d'un processus itératif avec le comité consultatif.
Étape 8	Peaufiner la compréhension de l'offre de services par des entrevues auprès de huit à dix intervenants terrain de chacune des trois régions (intervenants ayant au moins deux ans d'expérience dans la région et auprès de personnes avec des troubles mentaux graves).
Ces entrevues ont permis d'établir les forces et les défis de l'offre de services pour quatre personnes fictives aux besoins divers : 1) une personne en début de maladie et avec un besoin de réadaptation intensive et précoce ; 2) une personne difficilement en lien avec le réseau de la santé et nécessitant un suivi de réadaptation dans la communauté ; 3) une personne dont l'état mental est stable et qui présente un besoin de réinsertion sociale ou professionnelle ; 4) une personne âgée de 65 ans et plus avec des besoins multiples de réadaptation. *Questions types : Pouvez-vous me décrire le parcours typique que suivrait cette personne pour obtenir des services de réadaptation dans votre région ? Quelles sont les pratiques que vous voudriez voir disponibles dans votre région pour cette clientèle ? Quels sont les défis reliés* à l'offre de réadaptation pour cette clientèle *dans votre région ?*	

(*suite*)

Tableau 26.1.
Étapes réalisées dans le projet illustré (*suite*)

ÉVALUATION (*suite*)	
Étape 9	Définir les écarts et les enjeux pour chacune des régions en étroite collaboration avec les comités consultatifs (analyse et présentation de tableaux et de figures synthèses par l'équipe de recherche).
Étape 10	Analyser les facteurs qui soutiennent ou limitent l'implantation et l'intégration des meilleures pratiques de réadaptation dans ces organisations (discussion avec les comités consultatifs ; présentation de validation à d'autres partenaires hors du comité consultatif).
Étape 11	Soutenir les comités consultatifs pour qu'ils classent et priorisent les écarts et les enjeux et entament une réflexion en vue du plan d'action.
Étape 12	Soutenir le choix des actions à entreprendre pour l'amélioration des services offerts et pour répondre aux enjeux prioritaires définis par les détenteurs d'enjeux de chacune des trois régions.
Étape 13	Mise en commun des résultats des trois régions et définition d'enjeux et de stratégies d'action provinciales (discussion commune avec les comités consultatifs et autres partenaires provinciaux ou nationaux lors d'une journée provinciale).
DISSÉMINATION	
Étape 14	Diffuser les résultats en étroite collaboration avec les comités consultatifs (organisation de journées destinées aux intervenants, présentations à différentes instances décisionnelles, etc.).
Étape 15	Soutenir les comités consultatifs et autres partenaires impliqués pour mettre en œuvre les actions choisies (p. ex., coordonner des réunions communes, les mettre en contact avec d'autres partenaires ayant les mêmes préoccupations).

1. La construction de modèles logiques consiste à rendre explicites les ressources, les activités et les objectifs d'un programme dans le but de porter un jugement sur sa cohérence (Contandriopoulos *et al*., 2000). Selon Rossi, Freeman et Lipsey (1999), le modèle logique contient : 1) le plan d'utilisation des services (*service utilization plan*), c'est-à-dire la séquence logique des services et pratiques offerts dans le programme, et 2) le plan organisationnel (*program's organizational plan*), c'est-à-dire les ressources déployées et les processus d'utilisation des activités qui permettent d'offrir les services ciblés et d'atteindre les résultats escomptés.

Dans ce processus évaluatif, l'équipe de recherche, composée de quatre chercheurs, d'un coordonnateur et de plusieurs étudiants, avait comme principale tâche la collecte et l'analyse préliminaire des données (première modélisation logique ; passation des entrevues ; analyse des données d'entrevues ; analyse des documents fournis par les membres des comités consultatifs et des comptes rendus des rencontres ; conception de tableaux et de figures synthèses, etc.). L'analyse finale des données, leur interprétation, la diffusion des résultats ainsi que les décisions qui en ont découlé pour le plan d'action se sont faites avec les comités consultatifs. Lors de chacune des rencontres, le chercheur principal et le coordonnateur du projet présentaient sous différents formats les résultats préliminaires issus de chacune des étapes. Les comités consultatifs avaient alors comme rôle de compléter l'information, de valider les conclusions émises, de

confirmer ces conclusions auprès de leurs équipes respectives et de prendre les décisions requises pour les étapes à venir dans le projet (p. ex., continuer la collecte de données pour plus de raffinement, schématiser différemment les données du projet). Parmi les façons utilisées par l'équipe de recherche pour présenter les données aux comités consultatifs, on retrouvait : une carte géographique grand format de toutes les organisations impliquées, des schémas des trajectoires de services, un tableau complet détaillé de l'offre de services (organisations, services, ressources, activités), un tableau comparatif avec la littérature scientifique, des tableaux synthèses des enjeux et des forces. Ces documents papier étaient imprimés pour chaque membre des comités consultatifs en format régulier ou en format affiche, et chacun pouvait ainsi les commenter, les annoter et même les faire circuler dans son milieu pour validation. L'équipe de recherche était constamment préoccupée par la congruence des résultats avec la réalité du terrain (pertinence sociale) et plusieurs mécanismes de consultation et validation ont été mis en place pour aller dans ce sens (rencontres fréquentes avec les comités consultatifs, diffusion des données préliminaires, impression d'affiches et de tableaux pour diffusion dans les milieux, tenue d'un journal détaillé de recherche, etc.). Le coordonnateur de recherche tenait un journal pour chaque région à l'étude incluant l'ensemble des informations concernant les échanges avec les différentes partenaires (par courriel, par téléphone ou en personne), les présences aux rencontres, les comptes rendus des discussions (qui étaient enregistrées pour en faciliter la synthèse), ainsi que toutes informations ou impressions jugées pertinentes pour l'analyse et l'avancement du projet.

Les décisions au sein des comités consultatifs étaient majoritairement prises par consensus. Seule l'étape de priorisation des enjeux et des stratégies d'action a nécessité une démarche de technique nominale (voir section 2.4 pour plus de détails à ce sujet). Pour ce qui est du processus de diffusion des résultats et de mise en action, pour lequel des fonds supplémentaires de dissémination ont été reçus, l'équipe de recherche a continué à assumer un rôle de coordination des rencontres, mais en déplaçant clairement la direction du processus aux membres des comités consultatifs et instances décisionnelles présentes dans les régions (p. ex., Comité des réseaux locaux de services en santé mentale de l'Est de Montréal). La dernière année du projet a donc servi de transition pour que les gestionnaires membres des comités consultatifs s'approprient l'ensemble des résultats et passent à l'action (p. ex., en assumant la coordination et l'animation des rencontres, en présentant eux-mêmes les résultats à d'autres partenaires, en ciblant les projets à développer et les partenaires à impliquer). En parallèle, les résultats ont aussi fait l'objet d'une journée provinciale en présence des trois régions de l'étude ainsi que de webinaires avec les partenaires des autres régions du Québec, les associations provinciales concernées par le sujet et quelques

partenaires canadiens. Ces rencontres, animées par l'équipe de recherche et impliquant plusieurs partenaires provinciaux ou membres des comités consultatifs, ont permis de présenter la mise en commun des résultats issus des trois régions à l'étude, de généraliser la réflexion à un niveau provincial (et même hors du Québec) et de soutenir le réseautage autour d'enjeux communs (c.-à-d. ceux qui ont été soulevés dans les trois régions).

2.3. Résultats

Les résultats qui découlent de ce projet sont multiples. Ils ont d'abord concerné chaque région à l'étude et, ensuite, une mise en commun des résultats a été effectuée pour dégager une vision provinciale. Cette mise en commun a permis d'avoir une vue d'ensemble des forces, des défis et enjeux, et des stratégies d'action à mettre en œuvre pour l'amélioration des services de réadaptation en santé mentale au Québec. Les résultats sont le fruit d'une analyse conjointe entre l'équipe de recherche et l'ensemble des parties prenantes impliquées tout au long du projet. Aux fins de l'illustration, seulement quelques résultats (principalement ceux à portée provinciale) seront présentés ci-dessous.

2.3.1. Forces et défis de l'implantation des meilleures pratiques de réadaptation

Pour chacune des régions à l'étude, les analyses de cas ont permis de faire émerger les forces et les défis de l'implantation des meilleures pratiques de réadaptation associés à leur contexte respectif. Le tableau 26.2 présente ces résultats et permet de rendre compte des défis propres à chaque région ainsi que de ceux qui sont communs et dont la portée est provinciale. Ces forces et défis ont émergé à la suite de l'analyse des écarts entre l'offre réelle de services et la gamme des meilleures pratiques de réadaptation recommandée dans la littérature scientifique[5].

5. Selon la littérature scientifique, toute une gamme d'approches et de pratiques de réadaptation pour la clientèle ciblée sont recommandées (Dixon et al., 2010; Jung et Newton, 2009; NICE, 2009; SAMHSA, 2011). Pour connaître la totalité des approches et pratiques recommandées, les lecteurs peuvent consulter le site Web du projet: <http://www.iusmm.ca/le-cerris/projets-de-recherche/enjeux-et-readaptation.html> ou le site <http://cerrisweb.com>.

Tableau 26.2.
Forces et défis associés aux contextes des régions à l'étude

Contexte rural	Contexte mixte	Contexte urbain
Région type : Abitibi-Témiscamingue	Région type : Lanaudière	Région type : Est de Montréal
Forces ▪ Bonne variété de services (approx. 67 organisations impliquées) ▪ Services personnalisés et de proximité (soutien dans le milieu de vie, soutien aux études et à l'emploi, etc.) ▪ Services coordonnés et en continuité entre milieux publics et communautaires ▪ Grande collaboration entre milieux publics et communautaires	**Forces** ▪ Bonne variété de services (approx. 58 organisations impliquées) ▪ Accès à des approches spécialisées de réadaptation (soutien/suivi dans le milieu de vie, approches cognitivescomportementales, soutien en emploi, etc.) ▪ Plusieurs activités en intersectorialité pour le travail et le logement, notamment des tables de concertation ▪ Équipe mobile régionale de crise	**Forces** ▪ Grande variété de services tant dans les milieux publics que communautaires (approx. 89 organisations impliquées) ▪ Accès à des approches spécialisées de réadaptation (soutien/suivi dans le milieu de vie, soutien en emploi et aux études, interventions précoces pour jeunes en début de psychose, etc.) ▪ Plusieurs services destinés aux familles ▪ Plusieurs services d'intervention de crise dans la communauté
Défis ▪ Pas d'accès à des services de soutien de crise et de suivi intensif dans le milieu de vie ▪ Accès difficile sur l'ensemble du territoire aux approches spécialisées de réadaptation (approches cognitives-comportementales, soutien en emploi régulier, interventions précoces pour jeunes en début de psychose) ▪ Peu d'accès à des approches intégrées ▪ Méconnaissance des services destinés aux familles et aux proches	**Défis** ▪ Défi parfois de coordination et de continuité des services ▪ Représentation pas toujours commune des principes de réadaptation ▪ Accès difficile sur l'ensemble du territoire aux approches spécialisées de réadaptation (approches cognitives-comportementales, approches intégrées, interventions précoces pour jeunes en début de psychose)	**Défis** ▪ Difficulté à avoir une vue d'ensemble des services, à se retrouver dans la gamme de services offerts (mécanismes d'accès multiples) ▪ Méconnaissance entre les partenaires des pratiques et des services offerts ▪ Défi de coordination et de continuité des services ▪ Défi d'offrir une approche personnalisée, axée sur les besoins des personnes
Défis communs aux trois régions ▪ Trop peu d'implication des familles comme partenaires ▪ Peu d'accès à du soutien entre pairs et à des pairsaidants ▪ Pénurie de logements autonomes et manque de services résidentiels de type soutien en logement ▪ Pas assez de services destinés aux jeunes hors du milieu de la psychiatrie et en continuité avec la pédopsychiatrie		

2.3.2. Enjeux communs et stratégies d'action

Sept enjeux communs et plusieurs stratégies d'action ont émergé de la mise en commun des résultats des trois histoires de cas. Aux fins de l'illustration, une partie des résultats est présentée ici (c.-à-d. les trois enjeux prioritaires retenus par les comités consultatifs). Pour soutenir plus concrètement la mise au point de stratégies d'action, des pratiques innovatrices ciblées par l'équipe de recherche (en collaboration avec les comités consultatifs et autres partenaires) accompagnent la présentation de chaque enjeu commun. Ces pratiques innovatrices ont été présentées et ont servi d'exemples lors de la journée provinciale impliquant les trois régions à l'étude. Pour servir d'exemples, ces pratiques devaient être implantées au Québec (dans les régions à l'étude ou pas), amener une vision différente des services à offrir (en respectant les données de la littérature scientifique), répondre aux enjeux soulevés, ainsi que stimuler l'innovation et le réseautage entre les gestionnaires et les intervenants québécois. De plus, elles devaient avoir été suggérées ou approuvées par les comités consultatifs lors des rencontres de projet.

- 1er enjeu commun : augmenter l'accès et la coordination des services pour les jeunes

  S*tratégies d'action :* établir davantage de partenariats avec les milieux de l'éducation ; favoriser les activités de réadaptation et du soutien hors du milieu de la psychiatrie (p. ex., L'initiation Espace-Transition offert dans la communauté en mixité sociale <http ://www.etpsy.ca/>).

- 2e enjeu commun : mettre l'accent sur le soutien et les services dans la communauté hors du milieu de la psychiatrie

  Stratégies d'action : mettre à contribution les ressources de soutien dans la communauté hors du milieu de la psychiatrie (p. ex., Projet d'apprentissage alimentaire Cuisinons Ensemble offert par des formateurs-cuisiniers à l'organisme Le Murier <http ://www.lemurier.org/>) ; favoriser le développement de logements sociaux et le soutien en logement autonome (p. ex., Projet de logements sociaux en mixité sociale SHERPA offert dans la ville de Québec <http ://infopech.org/les-sept-volets-de-service/7-sherpa/>).

- 3e enjeu commun : encourager l'implication des familles et des proches et augmenter l'accès aux services pour eux

  Stratégies d'action : impliquer davantage les familles et les proches (avec le consentement de la personne ayant un trouble mental grave) dans les plans d'intervention et plans de crise (directives anticipées

[Ambrosini et Crocker, 2009]); soutenir le développement d'approches pour et par les familles plus orientées vers le développement d'habiletés plutôt qu'uniquement orientées vers le soutien (p. ex., Projet Apprendre à se rapprocher sans agressivité <http://www.iusmm.ca/recherche/projets-et-comites/apprendre-a-se-rapprocher-sans-agressivite.html>).

2.3.3. Plan de transfert des connaissances

Pour permettre la diffusion des résultats de la recherche et soutenir la mise en place de plans d'action, chaque comité consultatif a également mis en place un plan de transfert et d'échange des connaissances (avec le soutien de l'équipe de recherche et grâce à une subvention de dissémination). Diverses activités ont alors été proposées et organisées par les comités consultatifs. Certaines sont décrites ci-dessous :

- Journée de réseautage sous forme d'ateliers destinés aux intervenants et partenaires des réseaux locaux de services en santé mentale de l'Est de Montréal :
  - Discussion et partage autour des quatre histoires de cas cités plus haut.
  - Organisation et animation par le comité des réseaux locaux de services en santé mentale de l'Est de Montréal et l'équipe de recherche.
- Mise à jour des ressources sur le site Web de l'Agence de la santé et des services sociaux de l'Abitibi-Témiscamingue (ASSSAT) et conception d'une carte promotionnelle pour en favoriser la diffusion :
  - Collaboration entre l'ASSSAT, les réseaux locaux de services de santé mentale de l'Abitibi-Témiscamingue et l'équipe de recherche.
  - Embauche d'une agente de communication multimédia et travail avec un graphiste.
- Diffusion d'informations sur les meilleures pratiques pour l'ensemble des intervenants/utilisateurs de services/proches du Québec :
  - Association avec le Centre d'études sur la réadaptation, le rétablissement et l'insertion sociale (CERRIS), qui a un mandat de transfert et d'échange de connaissances au moyen des technologies de l'information (<http://www.cerrisweb.com>).

- Orientations et priorités définies en collaboration avec les comités consultatifs et autres partenaires provinciaux (consultation lors de la journée provinciale et des webinaires).

2.4. Réflexions sur les processus participatifs

Plusieurs aspects discutés dans cette section revisitent des enjeux décrits dans la première partie du chapitre. Tout au long du projet, l'équipe de recherche se devait de jouer plusieurs rôles : coordonnateur, chercheur responsable de la rigueur scientifique, animateur, facilitateur et vulgarisateur. L'un des principaux défis a été de maintenir une discussion ouverte la plus neutre possible et de maintenir un lien de confiance avec et entre les partenaires. À quelques reprises, principalement lors des discussions sur des enjeux, des stratégies d'action ou des représentations différentes des façons d'offrir les services de réadaptation, il y a eu un risque que les rencontres du comité consultatif deviennent un lieu de résistance au changement et de revendication plutôt qu'un lieu de rapprochement et d'amélioration des pratiques. Lors de ces occasions, la qualité de la démarche de coconstruction et d'action concertée a été influencée par les habiletés de facilitateur et de négociateur de membres de l'équipe de recherche. Par exemple, à quelques reprises et en particulier dans des situations de perceptions différentes des meilleures pratiques à valoriser, l'équipe de recherche (composée du chercheur principal et du coordonnateur de recherche) a dû discuter avec certains partenaires hors des rencontres du comité consultatif pour faire valoir l'importance de leur implication et de leur opinion dans le groupe et ainsi éviter leur retrait du comité consultatif.

D'autres enjeux ont pu créer à certains moments une confusion dans les attentes et dans les objectifs du projet. Par exemple, lors du passage à l'élaboration des stratégies d'action, certains partenaires se sont placés dans une position d'observateur passif plutôt que dans la position de membre actif, détenteur d'enjeux et de solutions. Les rôles attendus et remplis par l'équipe de recherche, les objectifs et les attentes envers les membres du comité consultatif ont alors dû être clarifiés et repositionnés, afin d'éviter ces situations et ainsi maintenir un espace pour le dialogue et la mise en action des idées émises. Pour certaines régions habituées à travailler ensemble, le processus a été beaucoup plus facile et porteur de solutions concrètes que pour d'autres. La présence autour de la table de près d'une vingtaine de partenaires a également soulevé le défi de la gestion des grands groupes. L'équipe de recherche devait constamment s'assurer de la représentativité des membres présents aux rencontres du comité consultatif afin qu'aucune organisation ne soit sur- ou sous-représentée à un moment ou à un autre

dans le projet. Les jeux de pouvoir et d'influence entre les partenaires ainsi que les difficultés parfois à discuter ouvertement des problèmes ont pu nuire au processus de négociation et de concertation. Dans les trois régions, l'étape de priorisation des enjeux et de mise au point des actions à entreprendre a d'ailleurs été plus longue que prévu, ce qui a nécessité d'ajouter une à deux rencontres avec les comités consultatifs. Les personnes présentes avaient de la difficulté à atteindre un consensus et à prioriser les stratégies à mettre au point. À la lumière de cette observation, l'équipe de recherche a utilisé la technique de groupe nominal (voir le chapitre 12 du présent ouvrage) pour soutenir cette étape, accélérer le processus de décision et permettre au groupe d'avancer vers un plan d'action concret (Delbecq, Van de Ven et Gustafson, 1975).

Le projet, qui intégrait comme méthodes l'analyse logique des programmations de réadaptation à l'analyse de cas multiples, a été une démarche originale de la recherche évaluative pour répondre aux défis d'application des connaissances et d'amélioration des pratiques de réadaptation en santé mentale au Québec. En établissant des liens entre la programmation de réadaptation présentement offerte dans les régions québécoises, les enjeux et les attentes sur la façon dont les services devraient être conçus tout en consultant la littérature scientifique, les acteurs concernés ont entamé un processus de réflexion exhaustif. D'ailleurs, ce processus a permis aux acteurs détenteurs d'enjeux de prendre conscience des écarts, des forces et des défis de leur région respective pour orienter leurs réflexions communes vers des actions à entreprendre (théorie de l'action (Patton, 1997]). On comprend alors qu'un processus d'amélioration des pratiques et des services ne peut se faire hors d'un processus participatif, car les acteurs concernés connaissent les enjeux de leur réalité terrain. Ils sont aussi en mesure de choisir les solutions et les stratégies d'action à mettre au point, et ce, dans le respect de la culture du milieu et des enjeux de leur contexte respectif.

Dans ce projet, les membres des comités consultatifs ont apprécié l'expérience d'une approche participative. Un bref questionnaire maison a permis d'apprendre que le projet a influencé ou influencera dans la prochaine année plusieurs de leurs décisions concernant : 1) l'offre de services de réadaptation (32 % : oui certaines décisions déjà ; 47 % : à venir dans la prochaine année) ; 2) les formations à offrir (21 % : oui certaines décisions déjà ; 58 % : à venir dans la prochaine année) ; 3) les comités de travail à mettre en place (42 % : oui certaines décisions déjà ; 37 % : à venir dans la prochaine année). Voici quelques extraits des commentaires écrits concernant les aspects que les membres des comités consultatifs ont le plus retenus de leur participation au projet.

La communication transparente – la volonté d'actualiser et de développer des modèles de partenariat pour offrir un meilleur service à la clientèle.

Élargissement des connaissances – réseautage – constat des enjeux de la région – belle structure pour «forcer» l'actualisation des priorités d'action.

L'importance du réseautage lorsqu'on est en région éloignée. Ce projet a un grand impact stimulant et j'espère qu'il contribuera à nous donner l'élan pour se «réorganiser».

L'importance de travailler dans le sens des meilleures pratiques en santé mentale.

CONCLUSION

La recherche participative est une approche de recherche où se rencontrent diverses parties prenantes, incluant des chercheurs et des partenaires du milieu, autour d'un objet de recherche qui les rallient. Cet objet de recherche émane de besoins provenant du milieu. Ce type d'approche et de contexte demande donc du temps pour cultiver des liens de confiance, une flexibilité par rapport aux méthodes de collecte de données et des qualités de vulgarisateur et de négociateur chez le chercheur et l'équipe de recherche. Pour ces derniers, la finalité de l'approche participative demeure la création de nouvelles connaissances qui doivent «refléter les résultats d'une rencontre entre le monde de la recherche et celui du terrain où la voix de l'ensemble des acteurs est entendue à travers le savoir produit» (Couture *et al.*, 2007, p. 218). Pour les acteurs impliqués dans l'approche participative, les retombées sont multiples. Elle offre aux détenteurs d'enjeux un espace de réflexion, un accompagnement structuré et rigoureux, et des moyens pour soutenir les changements de pratique.

RÉFÉRENCES

ABMA, T.A. (2000). «Stakeholder conflict: A case study», *Evaluation and Program Planning*, vol. 23, p. 199-210.

AMBROSINI, D.L. et A.G. CROCKER (2009). «Les directives psychiatriques anticipées (DPA) et le rôle de l'autonomie», *Santé mentale au Québec*, vol. 34, nº 2, p. 51-74.

ANADÓN, M. et C. COUTURE (2007). «La recherche participative, une préoccupation toujours vivace», dans M. Anadón (dir.), *La recherche participative. Multiples regards*, Québec, Presses de l'Université du Québec, p. 3-7.

ANADÓN, M. et L. SAVOIE-ZAJC (2007). «La recherche-action dans certains pays anglo-saxons et latino-américains», dans M. Anadón (dir.), *La recherche participative. Multiples regards*, Québec, Presses de l'Université du Québec, p. 11-30.

Anderson-Draper, M.-H. (2006). «Understanding cultural competence through the evaluation of “breaking the ice”: A project to generate critical knowledge about family violence within immigrant communities», *The Canadian Journal of Program Evaluation*, vol. 21, no 2, p. 59-79.

Baron, C. (2007). «Une investigation collaborative et développementale de l'expérience du pouvoir chez des gestionnaires postconventionnels», dans M. Anadón (dir.), *La recherche participative.Multiples regards*, Québec, Presses de l'Université du Québec, p. 125-157.

Brisolura, S. (1998). «The history of participatory evaluation and current debates in the field», *New Directions for Evaluation*, vol. 80, p. 25-41.

Cockburn, L. et B. Trentham (2002). «Participatory action research: Integrating community occupational therapy practice and research», *Canadian Journal of Occupational Therapy*, vol. 69, no 1, p. 20-30.

Contandriopoulos, A. (2003). «Inertie et changement», *Ruptures, revue transdisciplinaire en santé*, vol. 9, no 2, p. 4-31.

Contandriopoulos, A.P. Champagne F., Denis J.L., et Avargues M.C. (2000). «Evaluation in the health sector: Concepts and methods», *Revue d'Epidémiologie et de Santé publique*, vol. 48, no 6, p. 517-539.

Contandriopoulos, A. et Y. Souteyrand (dir.) (1996). *L'hôpital stratège. Dynamiques locales et offre de soins*, Paris, John Libbey Eurotext.

Cousins, J.B. et J.A. Chouinard (2012). *Participatory Evaluation Up Close*, Charlotte, Information Age Publishing.

Cousins, J.B. et C. Walker (2000). «Predictors of educators' valuing of systematic inquiry in schools», *Canadian Journal of Program Evaluation*, Édition spéciale, p. 25-52.

Cousins, J.B. et E. Withmore (1998). «Framing participatory evaluation», *New Directions for Evaluation*, vol. 80, p. 5-23.

Couture, C., N. Bednarz et S. Barry (2007). «Multiples regards sur la recherche participative. Une lecture transversale», dans M. Anadón (dir.), *La recherche participative. Multiples regards*, Québec, Presses de l'Université du Québec, p. 206-221.

Daigneault, P.M. et S. Jacob (2012). «Conceptualiser et mesurer la participation à l'évaluation», dans V. Ridde et C. Dagenais (dir.), *Approches et pratiques en évaluation de programmes*, Montréal, Les Presses de l'Université de Montréal, p. 233-254..

Daigneault, P.M., S. Jacob et J. Tremblay (2012). «Measuring stakeholder participation in evaluation: An empirical validation of the Participatory Evaluation Measurement Instrument (PEMI)», *Evaluation Review*, vol. 36, no 4, p. 243-271, <doi:10.1177/0193841X12458103>.

Delbecq, A., A.H. Van de Ven et D.H. Gustafson (1975). *Group Techniques for Program Planning: A Guide to Nominal Group and Delphi Processes*, Scott, Foresman.

Desgagné, S. (2007). «Le défi de coproduction de “savoir” en recherche collaborative», dans M. Anadón (dir.), *La recherche participative. Multiples regards*, Québec, Presses de l'Université du Québec, p. 89-121.

DESGAGNÉ, S. et N. BEDNARZ (2005). « Médiation entre recherche et pratique en éducation : faire de la recherche “avec” et plutôt que “sur” les praticiens », *Revue des sciences de l'éducation*, vol. 31, nº 2, p. 245-258.

DIXON, L.B, DICKERSON, F., BELLACK, A.S., BENNETT, M., DICKINSON, D., GOLDBERG, R.W., LEHMAN, A., TENHULA, W.N., CALMES, C., PASILLAS, R.M., PEER, J., et KREYENBUHL, J. (2010). « The 2009 Schizophrenia PORT Psychosocial Treatment Recommendations and Summary Statements », *Schizophrenia Bulletin* vol. 36, nº 1, p. 48-70.

DUBET, F. (1994). *Sociologie de l'expérience*, Paris, Seuil.

DUSSAULT, J., N. JAUVIN, M. VÉZINA et R. BOURBONNAIS (2012). *Preventing Violence Among Employees of the Same Work Organization. Evaluation of a Participatory Intervention*, Québec, Institut de recherche Robert-Sauvé en santé et en sécurité du travail (IRSST), Research REPORT R-739.

FALS BORDA, O. (1985). *Conocimiento y saber popular*, Bogota, Siglo XIX.

FREIRE, P. (1973). La concepción bancaria de la educación y la deshumanización. Dans P. Freire, H.Y. Fiori et J.L. Fiori (Eds). *Educación liberadora*, Bilbao, Zero-Zyx, p. 59-72.

GREEN, L.W., FRANKISH, C.J., GOLD, R., KREUTER, M.W., KREUTER, M.W, MERCER, S., MORISKY, D., NAIR, S., OTTOSON, J., POLAND, B., et ROOTMAN, I. (1995). *Guidelines and Categories for Classifying Participatory Research Projects in Health*, <http://lgreen.net/guidelines.html>, consulté le 23 janvier 2014.

GUBA, E.G. et Y.S. LINCOLN (1989). *Fourth Generation Evaluation*, Thousand Oaks, Sage Publications.

GUILLEMETTE, S. (2011). Étude de l'ajustement de pratiques vers une gestion différenciée de l'activité éducative par des directions d'établissement : expérimentation d'un modèle d'accompagnement collectif, Thèse de doctorat, Québec, Université de Sherbrooke.

HEALY, K. (2001). « Participatory action research and social work. A critical appraisal », *International Social Work*, vol. 44, nº 1, p. 93-105.

HERON, J. et P. REASON (1997). « A participatory inquiry paradigm », *Qualitative Inquiry*, vol. 3, nº 3, p. 274-294.

JUNG, X. et R. NEWTON (2009). « Cochrane Reviews of non-medication-based psychotherapeutic and other interventions for schizophrenia, psychosis, and bipolar disorder : A systematic literature review », *International Journal of Mental Health Nursing*, vol. 18, nº 4, p. 239-249.

KEMMIS, S. et R. MCTAGGART (2008). « Participatory action research/communicative action and the public sphere », dans N.K. Denzin et Y.S. Lincoln (dir.), *Strategies of Qualitative Inquiry*, 3e éd., Thousand Oaks, Sage Publications, p. 271-330.

MERCIER, C. (1997). « Participation in stakeholder-based evaluation : A case study », *Evaluation and Program Planning*, vol. 20, nº 4, p. 467-475.

MICHAUD, C. et P. BOURGAULT (2010). « Les devis de recherche non traditionnels », dans M.F. Fortin (dir.), *Fondements et étapes du processus de recherche*, Montréal, Chenelière Éducation, p. 362-371.

MORRISSETTE, J. (2012). « Quelques ficelles du métier de chercheur collaboratif », *Recherches qualitatives*, Hors Série, vol. 13, p. 5-19.

NATIONAL INSTITUTE FOR HEALTH AND CLINICAL EXCELLENCE – NICE (2009). *Schizophrenia : Core Interventions in the Treatment and Management of Schizophrenia in Adults in Primary and Secondary Care*, <http ://www.nice.org.uk/CG82>, consulté le 3 décembre 2010.

PATTON, M.Q. (1997). *Utilization-Focused Evaluation : The New Century Text*, 3e éd., Thousand Oaks, Sage Publications.

PATTON, M.Q. (2008). *Utilization-Focused Evaluation*, 4e éd., Thousand Oaks, Sage Publications.

PATTON, M.Q. et F. LABOSSIÈRE (2012). « L'évaluation axée sur l'utilisation », dans V. Ridde et C. Dagenais (dir.), *Approches et pratiques en évaluation de programmes*, Montréal, Les Presses de l'Université de Montréal, p. 145–160.

RIDDE, V. (2006). « Suggestions d'améliorations d'une cadre conceptuel de l'évaluation participative », *The Canadian Journal of Program Evaluation*, vol. 21, nº 2, p. 1-23.

RIDDE, V. (2009). « Une évaluation participative des processus avec renforcement du pouvoir d'agir de travailleurs de rue », dans V. Ridde et C. Dagenais (dir.), *Approches et pratiques en évaluation de programmes*, Montréal, Les Presses de l'Université de Montréal, p. 245-260..

ROSSI, P.H., H.E. FREEMAN et M.W. LIPSEY (1999). *Evaluation : A Systematic Approach*, 6e éd., Thousand Oaks, Sage Publications.

SAVOIE-ZAJC, L. (2012). « Du déroulement évolutif de la recherche-action au format linéaire l'écriture : quelques défis dans la rédaction et la diffusion de la recherche-action », *Recherches qualitatives*, Hors Série, vol. 13, p. 73-89.

SIMARD, P., M. O'NEILL, C.J. FRANKISH, A. GEORGE, M. DANIEL et M. DOYLE-WATERS (1997). *La recherche participative en promotion de la santé au Canada francophone*, Ottawa, Santé Canada.

STUFFLEBEAM, D.L. (2001). « Evaluation models », *New Directions for Evaluation*, vol. 89, p. 7-98.

SUBSTANCE ABUSE AND MENTAL HEALTH SERVICES – SAMHSA (2011). *SAMHSA's National Registry of Evidence-based Programs and Practices*, U.S. Department of Health and Human Services, <http ://www.nrepp. samhsa.gov/>, consulté le 11 avril 2011.

TORBERT, W.R. (1999). « The distinctive questions developmental action inquiry asks », *Management Learning*, vol. 30, nº 2, p. 189-206.

TREMBLAY, D.-G. (2005). « Les communautés de pratique : quels sont les facteurs de succès ? », *Revue internationale sur le travail et la société*, vol. 3, nº 2, p. 692-722.

WEAVER, L. et J.B. COUSINS (2005). « Unpacking the participatory process », *Journal of MultiDisciplinary Evaluation*, vol. 1, p. 19-40.

YIN, R.K. (2014). *Case Study Research. Design and Methods*, 5e éd., Thousand Oaks, Sage Publications.

NOTICES BIOGRAPHIQUES

VALÉRIE ALBERT est candidate au doctorat interdisciplinaire en santé et société avec spécialisation en ergonomie à l'Université du Québec à Montréal. Ses intérêts de recherche portent sur la réadaptation au travail et l'évaluation d'interventions en milieu de travail visant la prévention des troubles musculosquelettiques.

MARJOLAINE BEAUDRY est doctorante en psychologie industrielle et organisationnelle à l'Université du Québec à Montréal. Ses intérêts de recherche portent sur la qualité de la relation de travail entre les supérieurs immédiats et les employés ainsi que sur l'influence de celle-ci sur la santé psychologique au travail.

DJAMAL BERBICHE est statisticien sénior au Centre de recherche – Centre de santé et de services sociaux Champlain–Charles-Le Moyne, Université de Sherbrooke. Ses intérêts de recherche et d'application portent sur les modèles multiniveaux en général et de régression en particulier dans plusieurs domaines de la santé.

KARINE BERTRAND est professeure agrégée aux programmes d'études et de recherche en toxicomanie du Département des sciences de la santé communautaire de l'Université de Sherbrooke. Elle est chercheuse au Centre de recherche

du CSSS Champlain – Charles-Le Moyne, à l'Institut universitaire sur les dépendances et dans l'équipe de recherche en toxicomanie, le RISQ. Ses travaux de recherche portent sur la compréhension des trajectoires addictives et de services des personnes aux prises avec une consommation problématique de psychotropes, sur le développement et l'évaluation d'interventions psychosociales adaptées aux caractéristiques de divers sous-groupes, comme les adolescents et les personnes utilisatrices de drogues injectables.

MARIE-CHRISTINE BRAULT est sociologue et stagiaire postdoctorale à l'Institut de recherche en santé publique de l'Université de Montréal (IRSPUM). Ses intérêts de recherche portent sur les déterminants sociaux de la santé mentale et sur l'épidémiologie sociale et psychiatrique.

CATHERINE BRIAND est professeure agrégée à l'Université de Montréal (École de réadaptation) et chercheuse au Centre de recherche de l'Institut universitaire en santé mentale de Montréal. Ses intérêts de recherche portent sur l'analyse des facteurs qui favorisent et limitent l'implantation des meilleures pratiques de réadaptation en santé mentale et des services orientés vers le rétablissement dans le système de santé mentale québécois.

CHANTAL CARA est professeure titulaire et vice-doyenne aux études supérieures à la Faculté des sciences infirmières de l'Université de Montréal. Elle est chercheuse au sein du Réseau de recherche en interventions en sciences infirmières du Québec et au Centre de recherche interdisciplinaire en réadaptation. Ses intérêts de recherche portent sur les pratiques humanistes, la spiritualité et la phénoménologie.

MARC CORBIÈRE est professeur titulaire à l'École de réadaptation de l'Université de Sherbrooke. Il est aussi chercheur boursier sénior du Fonds de recherche du Québec – Santé (FRQS) au Centre de recherche du CSSS Champlain – Charles-Le Moyne. Ses intérêts de recherche portent sur les facteurs biopsychosociaux de la réinsertion et du retour au travail de personnes aux prises avec un trouble mental ou d'autres problèmes de santé. Il travaille aussi sur la conception et la validation d'outils de mesure pour ces populations ainsi que sur l'évaluation de programmes de réinsertion professionnelle (<http://www.retrame.ca>). Il est coauteur de l'ouvrage collectif *Du trouble mental à l'incapacité au travail* (Presses de l'Université du Québec, 2011).

DANIEL CÔTÉ est anthropologue de formation et chercheur à l'Institut de recherche Robert-Sauvé en santé et en sécurité du travail (IRSST). Il s'intéresse à l'étude des déterminants sociaux de l'incapacité de longue durée chez les personnes qui ont subi une lésion professionnelle. Ses travaux actuels portent sur le processus de réadaptation et de retour au travail dans le contexte de la rencontre interculturelle.

MARIE-FRANCE COUTU est professeure agrégée à l'École de réadaptation de l'Université de Sherbrooke, directrice du microprogramme en gestion d'invalidité et chercheuse au Centre de recherche du CSSS Champlain – Charles-Le Moyne. Elle est chercheuse boursière du Fonds de recherche du Québec – Santé (FRQS).

Elle est aussi psychologue dans le domaine de la santé depuis plus de dix ans et fait de la recherche appliquée dans le domaine de la réadaptation au travail, auprès de travailleurs ayant soit une douleur persistante d'origine musculosquelettique ou un problème de santé mentale qui les empêche d'avoir une vie saine et active au travail. Ses travaux sont effectués en collaboration avec les travailleurs, les milieux cliniques, les employeurs, les syndicats et les assureurs.

KATHY DAHL est ergothérapeute à l'Institut universitaire en santé mentale de Montréal. Elle est également étudiante à la maîtrise en sciences cliniques à l'Université de Sherbrooke et est chargée d'enseignement de clinique pour l'École de réadaptation de l'Université de Montréal. Elle s'intéresse plus particulièrement au domaine de la réadaptation en santé mentale et aux services pour les personnes présentant un trouble de la personnalité limite.

JULIE DESROSIERS est professeure adjointe de clinique à l'Université de Montréal au programme d'ergothérapie. Elle est actuellement candidate au doctorat en sciences cliniques à l'Université de Sherbrooke. Elle s'intéresse aux répercussions fonctionnelles des troubles mentaux et à la mesure du fonctionnement au quotidien.

ÉRIC DION est professeur à l'Université du Québec à Montréal au Département d'éducation et formation spécialisées. Il est directeur du Laboratoire sur les pratiques d'enseignement appuyées par la recherche. Il s'intéresse au développement de programmes de prévention des difficultés sur le plan du langage, de l'attention et de la lecture.

CLERMONT E. DIONNE est professeur titulaire au Département de réadaptation de la Faculté de médecine de l'Université Laval et chercheur sénior de l'axe Santé des populations et pratiques optimales en santé du Centre de recherche du Centre hospitalier universitaire de Québec. Ses travaux portent sur l'épidémiologie des douleurs musculosquelettiques.

PIERRE-ALEXANDRE DIONNE est candidat au doctorat en sciences cliniques de l'Université de Sherbrooke et au Centre de recherche du CSSS Champlain – Charles-Le Moyne. Ses intérêts de recherche portent sur l'utilisation de médicaments (pharmacoépidémiologie), l'étude des coûts de la maladie, l'évaluation économique des soins de santé et la pharmacoéconomie.

VÉRONIQUE DUPÉRÉ est professeure adjointe à l'École de psychoéducation de l'Université de Montréal. Elle est également chercheuse régulière à l'Institut de santé publique de l'Université de Montréal et au Centre jeunesse de Montréal – Institut universitaire. Elle fait aussi partie du Groupe de recherche sur les environnements scolaires. Ses intérêts de recherche portent sur les effets du milieu de vie sur le développement de l'enfant et de l'adolescent.

MARIE-JOSÉ DURAND est professeure titulaire à l'École de réadaptation de l'Université de Sherbrooke et titulaire de la Chaire de recherche en réadaptation au travail Bombardier/Pratt & Whitney. Elle est également directrice du CAPRIT (Centre d'action en prévention et réadaptation de l'incapacité au travail) qui vise à la

fois le développement de recherches appliquées et la formation continue pour les différents acteurs impliqués dans le retour au travail. Elle est chercheuse sénior au Centre de recherche CSSS Champlain – Charles-Le Moyne. Ses principaux intérêts de recherche sont le développement et l'évaluation d'outils de mesure et de programmes de réadaptation pour différentes populations ayant des incapacités au travail.

AMÉLIE FELX est conseillère clinique en réadaptation à l'Institut universitaire en santé mentale de Montréal (Direction du soutien à l'intégration sociale) et candidate au doctorat à l'Université de Montréal (Département de psychiatrie). Ses intérêts de recherche portent sur le travail conceptuel, l'évaluation des services de santé mentale et la mise en place de meilleures pratiques.

JORGE FLORES ARANDA est candidat au doctorat en sciences cliniques à l'Université de Sherbrooke. Il s'intéresse à la consommation de drogues chez les membres des minorités sexuelles, aux aspects méthodologiques de la recherche auprès des populations marginalisées ainsi qu'à la prévention des infections transmissibles sexuellement et par le sang (ITSS).

FRANCO FRACCAROLI est professeur de psychologie du travail et des organisations à l'Université de Trento (Italie). Il a été président de l'EAWOP (European Association of Work and Organizational Psychology) et doyen de la Faculté de science cognitive de l'Université de Trento. Ses intérêts de recherche portent sur l'insertion professionnelle et les facteurs organisationnels qui favorisent le maintien au travail pour les personnes avec un trouble mental. Il s'intéresse aussi au stress au travail et à l'analyse des risques psychosociaux dans les organisations.

FRANCES GALLAGHER est professeure agrégée à l'École des sciences infirmières de l'Université de Sherbrooke. Elle est chercheuse au Centre de recherche Étienne-Lebel du Centre hospitalier universitaire de Sherbrooke ainsi qu'au Centre affilié universitaire du Centre de santé et de services sociaux – Institut universitaire de gériatrie de Sherbrooke. Ses intérêts de recherche portent sur la prévention, la promotion de la santé, les soins en première ligne et l'*empowerment*. Elle a développé une expertise en recherche qualitative et sur les méthodes mixtes.

DANIELLE GRATTON est anthropologue et psychologue clinicienne à l'Hôpital juif de réadaptation à Laval (HJR) où elle occupe aussi la fonction de consultante en relations interculturelles. Elle est chercheuse au Laboratoire de recherche en relations interculturelles (LABRRI) de l'Université de Montréal. Elle est aussi directrice et fondatrice du Centre d'étude et d'intervention en relations interculturelles (CEIRI), une ressource conçue en vue de répondre aux besoins de formation et de consultation pour les membres du personnel, les gestionnaires et les décideurs des milieux communautaires, institutionnels, municipaux, juridiques et ministériels qui doivent développer une expertise dans les dynamiques pluriethniques.

MAUDE HÉBERT est professeure à l'Université du Québec à Trois-Rivières. Elle est infirmière et elle a réalisé ses études de cycles supérieurs auprès de femmes atteintes d'un cancer du sein. Elle enseigne la méthodologie de recherche et la relation d'aide. Elle s'intéresse au concept de transition entre la santé et la maladie chez des personnes atteintes d'un cancer ainsi qu'aux perceptions de la maladie.

NATHALIE HOULFORT est professeure à l'Université du Québec à Montréal en psychologie industrielle et organisationnelle et titulaire de la Chaire d'étude sur l'application des connaissances dans le domaine des jeunes et des familles en difficulté. Ses travaux examinent le rôle des processus motivationnels dans la vie quotidienne des travailleurs. Elle s'intéresse particulièrement aux facteurs organisationnels et individuels facilitant et contraignant le transfert des connaissances dans les organisations ainsi qu'aux conséquences du transfert. Ses autres projets de recherche actuels portent sur la conciliation travail/vie personnelle et sur la passion envers le travail.

MARIE-CLAUDE JACQUES est professeure à l'École des sciences infirmières de l'Université de Sherbrooke et chercheuse au Centre affilié universitaire du Centre de santé et de services sociaux – Institut universitaire de gériatrie de Sherbrooke. Elle enseigne les soins en santé mentale et en psychiatrie, ainsi que la méthode de la théorisation ancrée. Ses intérêts de recherche portent sur l'adaptation des personnes souffrant de troubles mentaux graves et vivant dans la précarité, ainsi que sur les meilleures pratiques visant cette clientèle.

MARY KANE est la présidente de Concept Systems, Inc., une société de recherche et de conseil qui répond aux besoins de chercheurs et d'évaluateurs dans divers domaines tels que les sciences sociales, la santé publique, la santé mentale et les politiques gouvernementales. Elle a un vif intérêt pour les méthodes qui augmentent la capacité des individus à participer à des recherches qui peuvent affecter leur vie et leurs communautés. Elle est coauteure du livre *Concept Mapping pour la planification et l'évaluation* (Sage Publications, 2007).

LISE LACHANCE est professeure titulaire au Département d'éducation et pédagogie de l'Université du Québec à Montréal (UQAM). Elle est chercheuse au Centre de recherche et d'intervention sur l'éducation et la vie au travail (CRIEVAT) de l'Université Laval et fait partie du Groupe de recherche et d'intervention sur la présence attentive (GRIPA) de l'UQAM ainsi que du Laboratoire sur l'adaptation personnelle, sociale et neuropsychologique (LAPERSONE) de l'Université du Québec à Chicoutimi. Ses travaux de recherche portent sur la conciliation des rôles de vie et le fonctionnement social. Ils visent à mieux comprendre le processus d'adaptation des individus lors de l'accomplissement de rôles multiples, d'importantes transitions ou d'événements de vie majeurs.

NATHALIE LANCTÔT est stagiaire postdoctorale à l'École de criminologie de l'Université de Montréal et au Centre d'étude sur le trauma de l'Institut universitaire en santé mentale de Montréal. Ses intérêts de recherche portent sur la violence en milieu de travail. Plus précisément, elle s'intéresse aux répercussions de la violence en milieu de travail, et aux déterminants du retour au travail et du maintien en emploi des travailleurs victimes ou témoins de tels actes.

NADINE LARIVIÈRE est professeure adjointe à l'École de réadaptation de l'Université de Sherbrooke. Elle est chercheuse au Centre affilié universitaire du Centre de santé et de services sociaux – Institut universitaire de gériatrie de Sherbrooke et chercheuse associée à l'Institut universitaire en santé mentale de Montréal. Ses intérêts de recherche principaux portent sur les répercussions fonctionnelles de vivre avec un trouble de la personnalité limite sur l'engagement dans des activités significatives, l'équilibre de vie et le sens des occupations ainsi que sur l'évaluation de programmes en santé mentale.

FRANÇOIS-ALBERT LAURENT est candidat au doctorat à l'Université du Québec à Montréal, sous la supervision de Nathalie Houlfort. Ses intérêts de recherche portent sur la motivation, la passion et le fonctionnement optimal au travail.

TANIA LECOMTE est professeure titulaire au Département de psychologie de l'Université de Montréal et chercheuse boursière sénior du Fonds de recherche en santé du Québec au Centre de recherche de l'Institut universitaire en santé mentale de Montréal (CRIUSMM). Ses recherches et ses publications portent sur l'amélioration des traitements psychosociaux (dont la thérapie cognitive-comportementale pour la psychose) et des services offerts aux personnes souffrant de troubles mentaux graves. Elle s'intéresse aussi aux personnes présentant une première psychose (<http://lespoir.ca>).

ALAIN LESAGE est professeur titulaire au Département de psychiatrie de l'Université de Montréal, psychiatre et chercheur à l'Institut universitaire en santé mentale de Montréal. Il codirige le réseau québécois de recherche sur le suicide. Ses intérêts de recherche portent aussi sur les besoins de services des personnes souffrant de troubles mentaux graves et sur les perspectives populationnelles.

NADIA L'ESPÉRANCE est agente de planification, de programmation et de recherche au Centre de réadaptation en dépendance Domrémy-de-la-Mauricie/Centre-du-Québec et professeure associée au Département de psychoéducation de l'Université du Québec à Trois-Rivières. Elle a terminé un stage postdoctoral aux programmes d'études et de recherche en toxicomanie de l'Université de Sherbrooke. Ses intérêts de recherche portent sur l'amélioration des services auprès des jeunes, des jeunes parents et des personnes en situation de rupture sociale.

JULIE MÉNARD est professeure de psychologie organisationnelle à l'Université du Québec à Montréal et directrice du laboratoire de recherche LEREPIT. Elle effectue ses recherches dans le domaine de la psychologie de la santé occupationnelle. Dans le cadre de sa carrière, elle s'est d'abord penchée sur les comportements éthiques et antisociaux au travail et sur le rôle de l'authenticité au travail dans le fonctionnement psychologique des travailleurs. Ses travaux les plus récents portent sur le répit des travailleurs et le détachement psychologique du travail. Ses études actuelles intègrent les approches basées sur la pleine conscience *mindfulness*, l'acceptation et l'engagement (Acceptance and Commitment Therapy – ACT) en tant qu'outils de gestion du stress au travail auprès de divers types de travailleurs.